

Health Technology Assessment

Schriftenreihe des Deutschen Instituts für
Medizinische Dokumentation und Information
im Auftrag des Bundesministeriums für
Gesundheit und Soziale Sicherung

Reihenherausgeber:

Friedrich Wilhelm Schwartz

Ansgar Gerhardus

Johannes Köbberling

Heiner Raspe

J.-Matthias Graf von der Schulenburg

Band 34

Methodenmanual für „HTA- Schnellverfahren“

UND

Exemplarisches „Kurz-HTA“: Die Rolle der quantitativen Ultraschall- verfahren zur Ermittlung des Risikos für osteoporotische Frakturen

Herausgeber:

**Dr. med. Monika Ekkernkamp, Dr. med. Dagmar Lühmann,
Prof. Dr. med. Dr. phil. Heiner Raspe**

Institut für Sozialmedizin
Universitätsklinikum Schleswig-Holstein
Campus Lübeck



Asgard-Verlag · Sankt Augustin

In der Schriftenreihe des Deutschen Instituts für Medizinische Dokumentation und Information im Auftrag des Bundesministeriums für Gesundheit und Soziale Sicherung werden Forschungsergebnisse, Untersuchungen, Umfragen usw. als Diskussionsbeiträge veröffentlicht. Die Verantwortung für den Inhalt obliegt der jeweiligen Autorin bzw. dem jeweiligen Autor.

Bibliografische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Alle Rechte vorbehalten.

© 2003 by Asgard-Verlag Dr. Werner Hippe GmbH,

Einsteinstraße 10 · 53757 Sankt Augustin

Telefon (0 22 41) 31 64-0

Telefax (0 22 41) 31 64 36

Internet: www.asgard.de

E-Mail: Info@asgard.de

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmung und die Einspeicherung und Verarbeitung in elektronische Systeme.

Titel-Nummer 270034

ISBN 3-537-27034-8

Druck: ICS Kommunikations-Service GmbH, Bergisch Gladbach

Gedruckt auf säurefreiem, alterungsbeständigem und chlorfreiem Papier.

VORBEMERKUNGEN

Bei der vorliegenden Abhandlung handelt es sich um den Abschlussbericht eines HTA-Projektes, dessen Ziel die Entwicklung von methodischen Arbeitsanleitungen war, welche die zeitnahe (3 - 6 Monate) Erstellung einer systematischen Verfahrensbewertung ermöglichen sollen. Es geht hierbei vorerst um die sogenannte "Effektivitätsbewertung" - gesundheitsökonomische Bewertungen waren nicht Gegenstand des Projektes.

Die Bearbeitung des Themas "Die Rolle der quantitativen Ultraschallverfahren zur Ermittlung des Risikos für osteoporotische Frakturen" diente der Erprobung der beschriebenen Arbeitsschritte. Der vorliegende Bericht besteht somit aus zwei getrennt gegliederten Teilen, die aber immer wieder Bezug aufeinander nehmen:

- dem "Methodenmanual für HTA-Schnellverfahren" und
- dem Kurz-HTA-Bericht "Die Rolle der quantitativen Ultraschallverfahren zur Ermittlung des Risikos für osteoporotische Frakturen".

Keiner der Berichtsteile stellt eine eigenständige Publikation dar. Die Autoren weisen hiermit ausdrücklich auf einige Limitationen der vorliegenden Arbeit hin:

a) Durch die zeitliche und finanzielle Begrenzung des Projektes konnten einige, aus unserer Sicht essentielle Arbeitsschritte in der Methodenentwicklung noch nicht unternommen werden - die vorliegenden Vorschläge müssen daher noch als Interimsprodukte bzw. Diskussionsgrundlage angesehen werden.

b) Wichtigste noch ausstehende Aufgabe ist unseres Erachtens die Erbringung des Validitätsnachweises der hier vorgeschlagenen Methodik. Es ist zu prüfen, ob eine nach unseren Vorschlägen unternommene Verfahrensbewertung zu anderen Schlussfolgerungen kommt als ein vollständiges HTA-Verfahren nach Standardmethodik. Die Überprüfung sollte an mehreren unterschiedlichen Themen vorgenommen werden. Weiterhin sind die einzelnen Arbeitsschritte zu verfeinern und zu ergänzen.

c) Es wäre in Zusammenarbeit mit gesundheitsökonomischen Arbeitsgruppen zu prüfen, inwieweit die schnelle Effektivitätsbewertung um gesundheitsökonomische Aspekte erweitert werden könnte bzw. inwieweit die Ergebnisse eine sinnvolle Basis für ökonomische Bewertungen liefern könnten.

d) Die hier vorgenommene Methodenentwicklung lässt nur begrenzt Schlussfolgerungen auf Zeit- und Ressourcenbedarf unter Realbedingungen zu. Die Arbeit im Methodenprojekt war von Entwicklungs- und Evaluationsarbeiten geprägt. Der echte Zeitbedarf für ein Kurz-HTA nach der hier vorgeschlagenen Methodik lässt sich nur schätzen.

Lübeck, im November 2003

Die Autoren

INHALTSVERZEICHNIS

A Abstracts	1
A.1 Methodenmanual für HTA-Schnellverfahren.....	1
A.2 Exemplarisches "Kurz-HTA"	2
B Executive Summaries	5
B.1 Methodenmanual für HTA-Schnellverfahren.....	5
B.2 Exemplarisches "Kurz-HTA"	9
I.C Methodenmanual für HTA-Schnellverfahren	14
I.C Hauptdokument	14
I.C.1 Zielsetzung des Berichtes.....	14
I.C.2 Hintergrund.....	15
I.C.2.1 Was ist Health Technology Assessment (HTA)?	15
I.C.2.2 Was ist ein „rapid assessment“?	18
I.C.2.3 Notwendigkeit und Konzept eines Schnellverfahrens	22
I.C.2.4 Beschreibung der Arbeitsabläufe/Vorgehen	26
I.C.3 Vorschläge für Standardized Operating Procedures (SOPs)	31
I.C.3.1 SOP 1: Feinabstimmung der Reviewfragestellung	31
I.C.3.1.1 Hintergrund - Status Quo (Ausführliches HTA).....	31
I.C.3.1.2 Methoden: Modifikations- / Kürzungspotential	32
I.C.3.1.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt.....	33
I.C.3.1.4 Diskussion: Bewertung, eigene Erfahrungen.....	38
I.C.3.1.5 Schlussfolgerung: SOP 1	39
I.C.3.2 SOP 2: Literaturrecherche für HTA-Schnellverfahren.....	40
I.C.3.2.1 Hintergrund: Status Quo (Ausführliches HTA).....	40
I.C.3.2.2 Methoden: Modifikations- / Kürzungspotential	41
I.C.3.2.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt.....	46
I.C.3.2.4 Diskussion: Bewertung, eigene Erfahrungen.....	50
I.C.3.2.5 Schlussfolgerung: SOP 2	52
I.C.3.3 SOP 3: Informationsbewertung, Bewertungsinstrumente	53
I.C.3.3.1 Hintergrund: Status Quo (Ausführliches HTA).....	53
I.C.3.3.2 Methoden: Modifikations-/ Kürzungspotential	61
I.C.3.3.3 Diskussion: Bewertung, eigene Erfahrungen.....	73
I.C.3.3.4 Schlussfolgerung: SOP 3.....	74
I.C.3.4 SOP 4: Datenextraktion	74
I.C.3.4.1 Hintergrund: Status Quo (Ausführliches HTA).....	74
I.C.3.4.2 Methoden: Modifikations- / Kürzungspotential	78
I.C.3.4.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt.....	78
I.C.3.4.4 Diskussion: Bewertung, eigene Erfahrungen.....	80
I.C.3.4.5 Schlussfolgerung: SOP 4	80

I.C.3.5 SOP 5: Informationssynthese	81
I.C.3.5.1 Hintergrund: Status Quo (ausführliches HTA).....	81
I.C.3.5.2 Methoden: Modifikations- / Kürzungspotential	87
I.C.3.5.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt.....	87
I.C.3.5.4 Diskussion: Bewertung, eigene Erfahrungen.....	88
I.C.3.5.5 Schlussfolgerung: SOP 5.....	88
I.C.3.6 SOP 6: Schlussfolgerungen, Aktualisierungsbedarf	89
I.C.3.6.1 Hintergrund: Status Quo	89
I.C.3.6.2 Methoden: Modifizierungs-/ Kürzungspotential.....	91
I.C.3.6.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt.....	95
I.C.3.6.4 Diskussion: Bewertung, eigene Erfahrungen.....	96
I.C.3.6.5 Schlussfolgerung: SOP 6.....	96
I.C.3.7 SOP 7: Dokumentationspflicht	98
I.C.3.7.1 Hintergrund: Status Quo	98
I.C.3.7.2 Methoden: Modifizierungs- / Kürzungspotential.....	98
I.C.3.7.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt.....	99
I.C.3.7.4 Diskussion: Bewertung, eigene Erfahrungen.....	99
I.C.3.7.5 Schlussfolgerung: SOP 7.....	100
I.D Literatur.....	103

II.C Exemplarisches "Kurz-HTA": Die Rolle der quantitativen Ultraschallverfahren zur Ermittlung des Risikos für osteoporotische Frakturen.....	108
II.C Hauptdokument	108
II.C.1 Policy Question / Ziel des HTA-Berichtes.....	108
II.C.2 Hintergrund.....	110
II.C.2.1 Hintergrund/Beschreibung des Krankheitsbildes	110
II.C.2.1.1 Definitionen und Begriffsbestimmungen	110
II.C.2.1.2 Frakturhäufigkeit und Osteoporoseassoziation.....	112
II.C.2.1.3 Relation von Knochendichte und Frakturrisiko.....	113
II.C.2.1.4 Präventions- und Behandlungsmöglichkeiten.....	114
II.C.2.2 Hintergrund/Beschreibung der Technologie.....	115
II.C.2.2.1 Einführung.....	115
II.C.2.2.2 Messmethoden	116
II.C.2.2.3 Genauigkeit, Richtigkeit und Präzision	120
II.C.2.2.4 Technologiegebrauch in Deutschland.....	127
II.C.3 Forschungsfragen.....	128

II.C.4 Methodik des HTA-Berichtes	129
II.C.4.1 Datenquellen und Recherchen.....	129
II.C.4.2 Bewertung der Informationen.....	129
II.C.5 Ergebnisse.....	131
II.C.5.1 HTA-Bericht: Homik & Hailey 1998	131
II.C.5.2 Systematische Literaturübersichten	135
II.C.5.3 Ergebnisse neuer Primärstudien.....	135
II.C.5.3.1 Gnudi et al. 2000.....	136
II.C.5.3.2 Stewart et al. 1999	138
II.C.5.3.3 Dargent-Molina et al. 1999.....	140
II.C.5.3.4 Pluijm et al. 1999	142
II.C.5.3.5 Huang et al. 1998.....	144
II.C.5.4 Ergebniszusammenfassung.....	149
II.C.6 Diskussion der Forschungsfragen	151
II.C.7 Schlussfolgerungen	159
II.D Literatur	161
Anhang	168

TABELLENVERZEICHNIS

Tabelle 1: Charakteristika schneller HTA-Programme	21
Tabelle 2: Komponenten eines HTA-Verfahrens.....	24
Tabelle 3: Modifikationsvorschläge aus der Expertenbefragung.....	37
Tabelle 4: Ergebnisse der Literaturrecherche in der Cochrane Library.....	48
Tabelle 5: Relevanzkriterien für die Literaturselektion	49
Tabelle 6: Evidenzhierarchie für Therapiestudien nach CRD	54
Tabelle 7: Evidenzhierarchie für Prognosefragestellungen.....	55
Tabelle 8: Phaseneinteilung der Diagnoseevaluierung.....	56
Tabelle 9: Stufenplan 1	65
Tabelle 10: Stufenplan 2	69
Tabelle 11: Endergebnisse der Reliabilitätstestung; Stufenpläne 1 und 2.....	70
Tabelle 12: Einzelauswertung der Reliabilitätstestung; Stufenpläne 1 und 2.....	72
Tabelle 13: Methodische Ausschlussgründe	73
Tabelle 14: Schlüsselkomponenten für die Datenextraktion	76
Tabelle 15: Extrahierte Daten für das Kurz-HTA "Ultraschall"	79
Tabelle 16: Glossar der Informationssynthese.....	82
Tabelle 17: Grading System von SIGN, modifiziert nach AHCPR	90
Tabelle 18: Effektmaße zur Beschreibung von dichotomen Studienergebnissen	93
Tabelle 19: Effektmaße zur Beschreibung der Ergebnisse diagnostischer Studien.....	94
Tabelle 20: Dokumentationsvorschriften nach "German Scientific Working Group"	98
Tabelle 21: Dokumentationsvorschläge für Rapid-HTA	98
Tabelle 22: Dokumentationsformular: Datenbank.....	100
Tabelle 23: Dokumentationsformular: Recherche	100
Tabelle 24: Dokumentationsformular: Qualitätsbewertung	100
Tabelle 25: Inzidenzrate der Schenkelhalsfrakturen in Deutschland im internationalen Vergleich.....	112
Tabelle 26: Messcharakteristika der quantitativen Ultraschallverfahren	120
Tabelle 27: QUS-Verfahren und Geräte.....	122
Tabelle 28: Korrelationskoeffizienten von BMD und QUS.....	125
Tabelle 29: Relevanzkriterien für die Literaturselektion	129
Tabelle 30: RR für Refraktur in Abhängigkeit von QUS und BMD	139
Tabelle 31: ROC-Analyse QUS und BMD-Messungen.....	139
Tabelle 32: Sensitivitäten von BMD, QUS, Gang und Alter für Hüftfrakturen	141
Tabelle 33: Sensitivitäten von Kombination aus BMD, QUS, Gang und Alter.....	141
Tabelle 34: Vorhersagewerte (RR) für BUA und SOS	143
Tabelle 35: Eckdaten der eingeschlossenen Studien	146
Tabelle 36: Evidenzlevel nach CEBM (2001) für Prognosestudien.....	175
Tabelle 37: Ein- und Ausschlusskriterien: QUS	178
Tabelle 38: Dokumentationsformular: Datenbank, QUS	178

Tabelle 39: Dokumentationsformular: Recherche, Medline + Healthstar	178
Tabelle 40: Dokumentationsformular: Recherche, Cochrane Library	179
Tabelle 41: Dokumentationsformular: Qualitätsbewertung	182
Tabelle 42: Evidenzbewertung der eingeschlossenen Materialien	183

ABKÜRZUNGS- UND SYNONYMVERZEICHNIS

a. p.	Anterior posteriorer Strahlengang
Abk.	Abkürzung
Abs.	Absatz
ADSoS	amplitudenabhängige Schallgeschwindigkeit
AHCPR	Agency for Health Care Policy and Research
AHFMR	Alberta Heritage Foundation for Medical Research
AHRQ	Agency for Healthcare Research and Quality
ARIF	Agressive Research Intelligence Facility, Department of Public Health and Epidemiology, University of Birmingham
Art.	Artikel
AUC	Area under the curve (receiver operating characteristic curve)
BDM	Bone Density Measurement
BMC	Bone Mineral Content
BMD	Bone Mineral Density
BMDs	BMD measured at the spine
BUA	Broadband Ultrasound Attenuation
BUB	Broadband-Ultrasound-Backscatter
Bzw.	beziehungsweise
Ca.	zirka
CAHTA	Catalan Agency for Health Technology Assessment
CCOHTA	Canadian Coordinating Office for Health Technology Assessment
CCTR	Cochrane Controlled Trials Register
CDSR	Cochrane Database of Systematic Reviews
CEBM	Center for Evidence-based Medicine (Oxford, UK)
CI	Confidence interval
CONSORT	Consolidated Standards of Reporting Trials
CRD	Center for Reviews and Dissemination
CT	Computertomographie
CV	Coefficient of Variation/Variationskoeffizient
d. h.	das heißt
DARE	Database of Abstracts of Reviews of Effectiveness
dB	Dezibel
DEC	Development Experience Clearinghouse
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information
DPA	Dual Photon Absorptiometry
DR	Distaler Radius
DU	Distale Ulna

DXA	Dual X-Ray Absorptiometry
EbCM	Evidence-based clinical medicine
EBM	Einheitlicher Bewertungsmaßstab
EbM	Evidence-based Medicine
Ecri	Emergency Care Research Institute
EED	Economic evaluation database
EPIDOS	Epidemiology de l'Osteoporose
etc.	et cetera
EVOS	Europäische Studie zur vertebrealen Osteoporose
FDA	Food and Drug Administration
Fx	Fraktur
ggf.	gegebenenfalls
GKV	Gesetzliche Krankenversicherung
HOS	Hawaii Osteoporosis Study
HSTAT	Health Services/Technology Assessment Text
HTA	Health Technology Assessment
IGeL	Individuelle Gesundheitsleistung
INAHTA	International Network of Agencies for Health Technology Assessment
ISTAHC	International Society of Technology Assessment in Health Care
k. A.	keine Angaben
KBV	Kassenärztliche Bundesvereinigung
KI	Konfidenzintervall
KMD	Knochenmineraldichte
MDK	Medizinischer Dienst der Krankenversicherungen
MDS	Medizinischer Dienst der Spitzenverbände der gesetzlichen Krankenkassen
MeSh	Medical Subject Headings
MHH	Medizinische Hochschule Hannover
MHz	Megahertz
MOOSE	Statement on Meta-Analysis of Observational Studies in Epidemiology
NCCHTA	National Coordinating Center for Health Technology Assessment
NEED	NHS Economic Evaluation Database
NHS EED	National Health Service Economic Evaluation Database
NHS	National Health Service
NLM	National Library of Medicine
OR	Odds Ratio
P	Patella
pQCT	Peripheral Quantitative Computed Tomography
QCT	Quantitative Computertomographie

QM	Qualitätsmanagment
QS	Qualitätssicherung
QUI	Quantitative Ultrasound Index
QUORUM	Quality of Reporting of Meta-Analysis of randomized controlled trials
QUS	Quantitative Ultrasound
RA	Radiographic Absorptiometry
RCT	Randomized controlled trial
ROC	Receiver Operating Characteristic
RR	Relatives Risiko
s.	siehe
SD	Standard Deviation = Standardabweichung
SOF	Study of Osteoporotic Fractures
SOP	Standard Operating Procedure
SOS	Speed of Sound/Schalleitungsgeschwindigkeit
SPA	Single Photon Absorptiometry
SXA	Single X-ray Absorptiometry
Tab.	Tabelle
Trent	NHS Trent Regional Office, Sheffield (UK)
T-Score	Abweichung des Messwertes vom Mittelwert des Referenzkollektivs junger Frauen (30 jährig)
u. a.	unter anderem
USA	United States of America
usw.	und so weiter
UTV	Ultrasound transmission velocity
vergl.	vergleiche
VOS	Velocity of Sound
WS	World Health Organisation
z. B.	zum Beispiel
z. Zt.	zur Zeit
Z-Score	Abweichung des Messwertes vom Mittelwert von gleichaltrigen und gleichgeschlechtlichen Gesunden

SYNONYME

Da sich im deutschsprachigen Raum noch keine HTA-Terminologie durchsetzen konnte, werden im vorliegenden Bericht folgende Begriffe synonym verwendet:

Für das Verfahren: HTA-Schnellverfahren
 Schnell-HTA
 Rapid Assessment

Für das Produkt: Kurz-HTA
 Kurz-HTA-Bericht
 Kurzgutachten

(- hierbei bezieht sich "kurz" auf den kurzen Erstellungszeitraum, nicht auf den Umfang des Werkes -)

SYMBOLLE

≈ ungefähr
∅ kein/keine
& und
? nicht bekannt

English Abstract: Method manual for rapid assessments

D Lühmann, M Ekkernkamp, H Raspe

Objective: HTA information required by decision makers needs to be both valid and presented in due time. In most HTA programs validity of information is achieved by applying systematic scientific methodologies in preparing the reports – procedures which are time- and / or resource consuming. Therefore our project aimed at the construction of standardized operating procedures (SOPs) for rapid HTAs which narrow the focus of the assessment and give up some of the depth of methodological information processing while maintaining the systematic approach and transparency.

Methods: On the basis of the methodology used in the German Scientific HTA Working Group, the HTA procedure was described in width (indicating the components an assessment could possibly comprise: e.g. background information, safety, efficacy, effectiveness, efficiency, utilization status, social-, ethical- or legal implication) and in depth (naming for each component the methodological approaches to gather adequate information). In order to construct a rapid assessment methodology, key elements for HTA and the essential methodological procedures were identified and formulated in terms of standardized operating procedures. The SOPs were tested and modified in the process of assessing the role of ultrasound devices for measuring bone density and -quality in osteoporosis.

Results: SOPs were constructed for seven key procedures in the course of performing a technology assessment:

SOP 1: Focussing the review questions in a modified Delphi procedure involving potential users as well as technological and methodological experts.

SOP 2: Construction of a hierarchical search strategy proceeding from high to low quality evidence.

SOP 3: Critical appraisal of information by use of short assessment forms.

SOP 4: Extraction of relevant information.

SOP 5: Guideline for choosing the adequate method of information synthesis.

SOP 6: Systematically deriving evidence-based conclusions.

SOP 7: Documentation of procedures and materials.

Conclusion: Preparation of an HTA-report according to the 7 SOPs described above, should not take longer than 3 to 4 months (calculated for one scientist trained in HTA-methodology and / or supported by an experienced HTA-team). The validity of the procedure as a whole remains to be established against the gold standard of the accepted review methodology.

English Abstract: Sample Rapid Assessment: The role of ultrasound devices for estimating osteoporotic fracture risk

M Ekkernkamp, D Lühmann, H Raspe

Background and Objectives: Today, the following definition of osteoporosis is widely accepted: "A disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to enhanced bone fragility and a consequent increase in fracture risk." So far bone density measurement by X-ray devices (e. g. DXA) has been the only non-invasive procedure for the estimation of bone fragility and associated intrinsic fracture risk. Against this background estimating fracture risk by quantitative ultrasound (QUS) measurement seems more attractive because the procedures do not apply ionizing radiation and are less costly. While the devices are being introduced into the (private) health care market, the following assessment was performed to clarify their potential benefit in the context of osteoporosis management with a special emphasis on the estimation of fracture risk.

Methods: The HTA has been performed according to the methodology suggested in SOPs 1 to 7 in part one of this report.

Five key questions were identified on the basis of a modified delphi procedure:

1. Compared to DXA, how valid are fracture risk estimates by quantitative ultrasound?
2. Which of the different ultrasound procedures yields the most valid results?
3. Before introducing the procedure into clinical practice: what measures have to be taken for quality assurance?
4. What kind of research is needed to further clarify the role of quantitative ultrasound measurements in the prevention of osteoporotic fractures?
5. How can the answers to questions 1 to 3 be used to inform consumers?

A systematic literature search found one HTA-report and five primary studies that matched the inclusion criteria and satisfied methodological requirements.

Results and Discussion of the key questions: 1. Data from prospective trials show that estimates of fracture risk for periods from six months to five and a half years are almost identical. Results of a limited number of studies suggest that a combination of QUS and DXA results possibly increases the predictive value of measurements. At the moment the available evidence does not justify recommendations as to which measurement procedure at which location should be used to predict fractures.

2. The usefulness of QUS measurements depends on their validity (accuracy and precision) and on the size and reliability of the risk estimates. Taking into considera-

tion validity, prognostic value and availability of devices transversal measurements of Broadband ultrasound attenuation (BUA) and Speed of Sound (SOS) at the calcaneus seem to be feasible.

3. The availability of reference data from a large healthy European population should be one of the prerequisites for the introduction of QUS measurements into clinical practice in Germany. Furthermore all staff performing measurements and interpretation of results should receive standardized training in a specialized center of competence (certificate!).

4. Acknowledging the uncertainties concerning the usefulness of QUS in everyday clinical practice further research, basic as well as clinical, seems necessary. Especially interventional trials are needed.

5. Communication with consumers should emphasize the developmental status of QUS for fracture risk prediction: therapeutical consequences of low QUS results are largely unclear and furthermore the procedure has not yet proven its usefulness for monitoring disease progress.

Conclusions: As long as the substrate of the different QUS measurements is not exactly known the role of QUS in the management of osteoporosis cannot be clarified. For DXA the use of Z- and T-Scores allows distinction of different disease categories but since these categories do not apply to ultrasound measurements this further limits the usefulness of the procedure.

At this stage of research there is no clear indication for QUS measurements in the diagnostic process of osteoporosis or in screening for osteoporosis. Consumers and patients ought to be informed on this. In the future, estimating fracture risk by ultrasound measurement seems possible. However its current use should be limited to research context and adhere to strict quality assurance measures.

Patients and consumers who want to buy QUS measurements should be accurately informed on the developmental state of the technology.

DANKSAGUNG

Die Autoren bedanken sich bei folgenden Personen für wichtige und hilfreiche Anregungen, Rückmeldungen, Kritik und Verbesserungsvorschläge:

Dr. R. Barkmann, Dr. B. Gibis, Prof. C.C. Glüer, Prof. W. Knopp, Dr. S. Kupsch, PD Dr. M. Perleth, Prof. J. Windeler.

A Abstracts

A.1 Abstract: Methodenmanual für HTA-Schnellverfahren

D Lühmann, M Ekkernkamp, H Raspe

Hintergrund und Zielsetzung: Informationen aus systematischen Verfahrensbewertungen (HTA-Informationen), die als Grundlage für gesundheitspolitische Entscheidungen dienen sollen, müssen einerseits zuverlässig und vollständig sein, andererseits zeitnah präsentiert werden. Ziel des hier berichteten Projektes war die Entwicklung von standardisierten Vorgaben für die Durchführung und die Erstellung von Kurz-HTA-Berichten. Durch eine stark fokussierte Fragestellung erfolgt die Durchführung des Kurz-HTA methodisch teilweise weniger komplex als die Durchführung eines ausführlichen HTA, jedoch unter strikter Beibehaltung des systematischen Ansatzes und der absoluten Transparenz.

Methodik: Auf der Basis der Methodenvorgaben für die deutsche HTA-Arbeitsgruppe (Bitzer et al. 1998) wurde eine zweidimensionale Beschreibung des HTA-Prozesses nach Elementen und Bearbeitungsmethodik unternommen. Für die methodischen Arbeitsschritte eines "Kurz-HTAs" wurden nun Schlüsselbestandteile und essentielle methodische Arbeitsschritte zu ihrer Erstellung identifiziert und als SOPs formuliert. Die SOPs wurden entworfen, getestet und modifiziert im Rahmen der Bearbeitung des Themas "Die Rolle der quantitativen Ultraschallverfahren zur Ermittlung des Risikos für osteoporotische Frakturen". Die dabei gemachten Erfahrungen wurden für jede SOP einzeln diskutiert und bewertet.

Ergebnisse: Für sieben Arbeitsschritte im Rahmen einer Verfahrensbewertung wurden SOPs vorgeschlagen:

SOP 1: Fokussierung der präzisen Fragestellung, SOP 2: Informationsrecherche, SOP 3: Kritische Literaturbewertung, SOP 4: Datenextraktion, SOP 5: Datensynthese, SOP 6: Systematische Ableitung von Schlussfolgerungen und Festschreibung des Aktualisierungsbedarfs und SOP 7: Dokumentationspflichten.

Diskussion (Hauptpunkte zusammengefasst): Das in SOP 1 vorgeschlagene Abstimmungsverfahren (modifiziertes, email-basiertes Delphi-Verfahren) stellte sich als sehr zeitaufwändig heraus. Eine Beschränkung des Projektbeirates zahlenmäßig auf höchstens vier Mitglieder, die verstärkt auf ihre Mitverantwortung für das Gesamtvorhaben hinzuweisen sind, erscheint sinnvoll.

Das für SOP 3 entwickelte Instrument zur Bewertung der Qualität von prognostischen Studien zeigte bei der Reliabilitätstestung keine zufriedenstellenden Resultate und muss vor weiterem Einsatz überarbeitet werden.

Schlussfolgerungen: Die Erstellung eines systematischen und transparenten Kurz-HTA unter Einsatz der oben skizzierten SOPs sollte innerhalb von drei bis vier Monaten möglich sein. Die Evaluation der Methode gegen einen Goldstandard steht noch aus.

A.2 Abstract: Exemplarisches "Kurz-HTA"

M Ekkernkamp, D Lühmann, H Raspe

Hintergrund: Die am weitesten verbreitete Definition der Osteoporose geht auf eine Konsensuskonferenz von 1991 zurück: "Eine Erkrankung, die durch niedrige Knochenmasse und eine Beeinträchtigung der Mikroarchitektur des Knochengewebes gekennzeichnet ist und so zu einer erhöhten Knochenbrüchigkeit und daraus folgend zu einem erhöhten Risiko führt, Frakturen zu erleiden". Bisher war die osteodensitometrische Messung der Knochenmineraldichte (KMD) mit Röntgenstrahlen das einzige nicht-invasive Verfahren zur Abschätzung der Knochenbrüchigkeit und damit des intrinsischen Frakturrisikos. Vor dem Hintergrund fehlender ionisierender Strahlung und vergleichsweise geringerer Kosten haben in den letzten Jahren unterschiedliche quantitative Ultraschall-Verfahren (QUS) zunehmende Verbreitung bei der Abschätzung des Risikos für „osteoporotische“ Frakturen gefunden. Vor dem Hintergrund der schnellen und zunehmenden Verbreitung dieser noch jungen Technologie in der Osteoporosediagnostik sollen deren Stellenwert und Wertigkeit in diesem HTA-Schnellverfahren untersucht werden. Dabei liegt der Schwerpunkt auf der Zusammenfassung der Ergebnisse der Frakturrisikoanalyse.

Methodik: Das exemplarische Kurzgutachten wurde systematisch nach den in Teil I dieses Berichtes vorgeschlagenen SOPs 1 - 7 erarbeitet.

Ergebnisse: Im modifizierten Delphi-Verfahren wurden folgende präzise Fragestellungen als vordringlich zu beantworten identifiziert (geordnet nach Priorität):

1. Wie valide lässt sich das Frakturrisiko mithilfe der quantitativen Ultraschall-Verfahren im Vergleich zum DXA-Verfahren abschätzen?
2. Welches der unterschiedlichen QUS-Verfahren ist hierzu am besten geeignet?
3. Welche qualitätssichernden Maßnahmen sind dann vor einem Einsatz in der ärztlichen Praxis zu realisieren?
4. Welchen weiteren Forschungsbedarf zur Klärung des Stellenwertes der Ultraschalldiagnostik im Rahmen der Frakturprävention gibt es?
5. Wie wären die unter 1 – 3 ermittelten Ergebnisse zur Patienteninformation und zum Verbraucherschutz zu nutzen?

Zur Beantwortung der Fragen konnte die Literaturrecherche / -selektion einen HTA-Bericht und fünf prospektive Einzelstudien mit zentraler Relevanz auffinden. Die Arbeiten entsprachen den per Checkliste abgefragten Qualitätskriterien.

Ergebnisse und Diskussion der Forschungsfragen: 1. Eine prognostische Gleichwertigkeit der Ultraschallsonometrie gegenüber DXA wurde belegt; QUS und DXA ermitteln, jeweils für sich betrachtet, vergleichbare relative Risiken für Frakturen innerhalb eines Beobachtungszeitraums von $\frac{1}{2}$ bis $5 \frac{1}{2}$ Jahren. Die Kombination von QUS und DXA bei multivariater Betrachtung führt möglicherweise zu einer verbesser-

ten Frakturprädiktion, die uns vorliegende Datenbasis zu dieser Frage ist allerdings begrenzt. Empfehlungen, welche Verfahren und Messorte zur Prädiktion welcher Frakturtypen herangezogen werden könnten, lassen sich aus diesen Daten allerdings noch nicht ableiten, hier besteht weiterer Forschungsbedarf.

2. Die Eignung von Ultraschallverfahren ist abhängig von der Zuverlässigkeit (Richtigkeit und Präzision) der Messungen und - in Abhängigkeit vom Messergebnis - der Höhe und Sicherheit des geschätzten Relativen Risikos für Frakturen. Berücksichtigt man Reliabilität, prognostischen Wert und Verbreitung gleichzeitig, dann spricht vieles für die transversale Messung von BUA/SOS am Kalkaneus.

3. Eine wichtige Voraussetzung für die Einführung einer quantitativen Ultraschallmethode zur Knochenuntersuchung in die klinische Praxis in Deutschland sollte das Vorhandensein eines genügend großen und repräsentativen Referenzkollektivs gesunder Frauen und Männer aus unserer europäischen Region sein. Größte Sorgfalt muss der Schulung der die Messung Durchführenden und Beurteilenden gewidmet werden. Die Ausbildung sollte unter standardisierten Bedingungen an Kompetenzzentren durchgeführt werden und dem Erwerb eines Befähigungsnachweises abschließen

4. Angesichts der vielfältigen Unklarheiten und Unsicherheiten im Bereich der QUS-Verfahren scheint eine Fortführung und auch Intensivierung der medizinischen und physikalischen Grundlagen- und krankheitsbezogenen Forschung erforderlich. Von großer Bedeutung scheinen darüber hinaus Interventionsstudien.

5. Auch Verbraucherinformationen sollen vermitteln, dass sich das Verfahren noch in der Entwicklung befindet, dass die therapeutischen Konsequenzen erniedrigter Befunde unklar sind und dass derzeit Verlaufskontrollen kaum möglich sind.

Schlussfolgerungen: Solange der Gegenstand der verschiedenen QUS-Messungen nicht klar ist, kann die Frage nach dem Beitrag des Verfahrens zur Osteoporosediagnose nicht beantwortet werden. Bei DXA-Messungen hat sich international der Gebrauch von T- und Z-Werten eingebürgert, um zu kategorialen Urteilen (z.B. „Osteopenie“, „Osteoporose“ nach WHO) zu kommen. Davon scheinen die QUS-Messungen noch weit entfernt. Es besteht dringender Bedarf zu weiterer medizinischer und physikalischer Forschung.

Es besteht zum jetzigen Zeitpunkt keine klare Indikation, Ultraschalluntersuchungstechniken in der Osteoporose-Diagnostik einzusetzen. Hierüber müssen Verbraucher und Patienten aufgeklärt werden.

Prinzipiell scheint ein zukünftiger Einsatz von QUS-Verfahren für die Frakturrisikoabschätzung möglich. Ihr Stellenwert für die Risikoerkennung und klinische Prävention ist derzeit aber noch unsicher. Der Gebrauch sollte daher an wissenschaftliche Fragestellungen und vor allem auch an eine konsequente Qualitätssicherung gebunden sein. Über Abweichungen hiervon sind diejenigen zu informieren, die das Verfahren, etwa als IGEL-Leistung, in Anspruch nehmen wollen. Die Datenlage für einen Ein-

satz der QUS-Verfahren in der Verlaufskontrolle ist begrenzt und kontrovers. Sichere Aussagen sind zur Zeit nicht möglich, auch weil die therapeutische Beeinflussung und Veränderungssensitivität der QUS-Parameter zu wenig geklärt sind. Einschränkend ist zu betonen, dass dieser Teilaspekt nicht Gegenstand unseres HTA-Schnellverfahrens war.

B Executive Summary

B.1 Methodenmanual für HTA-Schnellverfahren

D Lühmann, M Ekkernkamp, H Raspe

Hintergrund und Zielsetzung: Informationen aus systematischen Verfahrensbewertungen (HTA-Informationen), die als Grundlage für gesundheitspolitische Entscheidungen dienen sollen, müssen einerseits zuverlässig und vollständig sein, andererseits zeitnah präsentiert werden. Im Rahmen des deutschen HTA-Forschungsprojektes (finanziert durch das Bundesministerium für Gesundheit) wurde zur Sicherstellung der Validität von Informationen eine systematische, wissenschaftlich-methodische Vorgehensweise für die Erstellung von Verfahrensbewertungen zusammengestellt und für die deutsche HTA-Arbeitsgruppe als verbindliche Arbeitsanweisung festgelegt (s. Bitzer et al. 1998). Nach dieser zeit- und ressourcenintensiven Arbeitsweise dauerte eine Themenbearbeitung mindestens 12-18 Monate und machte damit HTA als Instrument zur Politikberatung für viele dringliche Fragestellungen untauglich.

Methodik: Auf der Basis der oben angesprochenen Methodik der deutschen HTA-Arbeitsgruppe (German Scientific HTA Working Group) wurde eine zweidimensionale Beschreibung des HTA-Prozesses versucht: eine Dimension benennt mögliche Bestandteile eines HTAs (z. B. Hintergrundinformationen, Sicherheit, Wirksamkeit unter Studien- und Alltagsbedingungen, Kostenwirksamkeit, Technologiegebrauch, soziale, ethische oder juristische Implikationen), die zweite Dimension beschreibt für jeden Bestandteil die methodischen Herangehensweisen an die Informationsgewinnung und –synthese, geordnet nach Komplexität. Für die methodischen Arbeitsschritte eines "Kurz-HTAs" wurden nun Schlüsselbestandteile und essentielle methodische Arbeitsschritte zu ihrer Erstellung identifiziert und als SOPs formuliert. Die SOPs wurden entworfen, getestet und modifiziert im Rahmen der Bearbeitung des Themas "Die Rolle von Ultraschallverfahren zur Frakturrisikoabschätzung bei Osteoporose".

Ergebnisse: Für sieben Arbeitsschritte im Rahmen einer Verfahrensbewertung wurden SOPs vorgeschlagen:

SOP 1: Fokussierung der präzisen Fragestellung

Folgende Arbeitsschritte sind in SOP 1 zusammengefasst:

1. Die Aufarbeitung von Hintergrundinformationen (z. B. orientierende Literaturrecherchen nach Schlüsselquellen; Expertenkonsultationen; Herstellerinformationen; administrative Daten)
2. Bildung eines Projektbeirats (maximal vier Personen – s. u.)

3. Entwurf von "Policy Question" und exakten Fragestellungen für das HTA
4. Priorisierung und Konsentierung der zentralen Fragestellung im modifizierten, elektronischen Delphi-Verfahren.

SOP 2: Informationsrecherche

Zusammenstellung der besten verfügbaren Informationen zur Fragestellung, dabei soll eine hohe Präzision zur Zeitersparnis führen. Folgende Arbeitsschritte sind in SOP 2 enthalten:

1. Definition von Ein- und Ausschlusskriterien (inhaltlich und methodologisch)
2. Plan einer hierarchisch strukturierten Suchstrategie:
 - nach Publikationen mit systematischen Literatursynthesen: (HTA-Berichte, systematische Reviews)
 - Primärstudien, zur Abdeckung der "Lücke" zwischen Publikationsdatum der Reviews und Gegenwart
 - Primärstudien, wenn es keine systematischen Reviews gibt.
3. Auswahl von Datenbanken:
 - HTA-Berichte: CRD-HTA, INAHTA, ISTAHC, HSTAT
 - Systematische Reviews: CDSR, DARE, Medline (spezielle Suchstrategie)
 - Primärstudien: Medline, Zeitrahmen s. o.
 - Weitere Datenbanken für Spezialfragestellungen
4. Suchstrategie: hoch sensitiv, gefolgt von manueller Selektion
5. Bestellung: nach Ein- und Ausschlusskriterien
6. Bestandkonferenz: nach Abschluss der Recherchen (Telefon)konferenz zum weiteren Prozedere:
 - protokollkonform
 - Modifikation der Fragestellung
 - Abbruch des Projektes

SOP 3: Kritische Literaturbewertung

Zur Bewertung der methodischen Qualität von Reviews und prospektiven Kohortenstudien wurden aus Itempools von publizierten Bewertungsinstrumenten zwei Kurzinstrumente zusammengestellt, deren Fragen mit "ja" oder "nein" zu beantworten sind. Beide Instrumente erlauben eine stufenweise Bewertung der Publikationen, wobei ein "ja" auf die vorangegangene Frage Voraussetzung für die Beantwortung der Folgefrage bildet. Ein solches Vorgehen erlaubt einen Abbruch der Bewertungsprozedur, sobald eine Frage mit "nein" beantwortet werden muss. Ausnahmen sind zu begründen (z. B. wenn die vorliegende Publikation die einzige Studie zu einer bestimmten Fragestellung beschreibt).

Die kritische Bewertung von Reviews umfasst:

- die Relevanz für die HTA-Fragestellung (Patienten, Interventionen, Outcomes)
- die Informationsrecherche (Datenbanken, Strategien, andere Informationsquellen)

- die methodische Qualitätsbewertung (Ein-, Ausschlusskriterien; Methodik der Qualitätsbewertung)
- Art und Qualität der Informationssynthese
- Stringenz der Schlussfolgerungen

Das Bewertungsinstrument für prospektive Studien erfasst:

- Zusammenstellung und Beschreibung der Studienpopulation
- Validität und Vergleichbarkeit der Expositionsmessung
- Validität und Vergleichbarkeit der Zielgrößenmessung
- Länge und Vollständigkeit des Follow-Up
- Statistische Analyse (Adjustierung für prognostische Faktoren)

SOP 4: Datenextraktion

Dient der Vorbereitung der Informationssynthese durch:

1. Identifikation der Art der benötigten Informationen
 - zur Beantwortung der HTA-Fragestellung(en)
 - zur Erklärung von Heterogenität (Sensitivitätsanalyse)
2. Spezifikation der zu extrahierenden Daten
 - allgemeine Informationen (z. B. Autoren, Publikationsdatum, Studiendesign)
 - spezifische Informationen: Studiencharakteristika (Population, Intervention(en), Zielgrößen); Follow-Up Informationen, Ergebnisse, Kommentare
3. Konstruktion des elektronischen Datenextraktionsformulars

SOP 5: Datensynthese

1. Tabellarische Darstellung der Studien (obligat): methodologische Qualität, Population, Intervention(en), Outcomes, Setting(s), Ergebnisse (Art, Größe, Robustheit)
2. Beschreibung der Schlüsselergebnisse im Text
3. Entscheidung ob eine Meta-Analyse notwendig und machbar wäre:
 - ist eine Meta-Analyse notwendig um aussagekräftige Ergebnisse zu erhalten?
 - ist eine Meta-Analyse aus methodischer Sicht machbar?
 - Durchführung der Meta-Analyse

SOP 6: Systematische Ableitung von Schlussfolgerungen und Festschreibung des Aktualisierungsbedarfs.

Abhängig von der Menge an notwendigen Kommentaren können die Schlussfolgerungen in einer ausführlichen Diskussion oder einem kurzen Schlussabsatz gezogen werden. Sie enthalten in jedem Fall:

1. Antwort(en) auf die zentralen HTA-Fragestellungen (Menge und Qualität der verfügbaren Informationen; Ergebnisse in absoluten und relativen Messgrößen; Notwendigkeit für Extrapolationen)
2. Formulierung von Empfehlungen
3. Beschreibung von Forschungsbedarf
4. Benennung eines "Verfallsdatums" unter Berücksichtigung der zukünftigen Technologieentwicklung, Demographie und Nachfrageverhalten

SOP 7: Dokumentationspflichten

Für die Kurz-HTAs werden zwei separate Dokumentationsstrategien vorgeschlagen: die Kerndokumentation wird als Anhang zum HTA-Bericht publiziert, die ausführliche Dokumentation kann auf Nachfrage von den Autoren zur Verfügung gestellt werden. Sortiert nach SOPs sollen folgende HTA-Komponenten dokumentiert werden, um die Transparenz der Verfahrensbewertung sicher zu stellen:

	Kerndokumentation	Ausführliche Dokumentation
SOP 1	Mitglieder des Projektbeirates Delphi-Prozess: Fragestellungen vorher und nachher	Protokoll des Delphi-Verfahrens, Korrespondenz, anonymisierte Voten und Kommentare, Zeitrahmen
SOP 2	Ein- und Ausschlusskriterien, Datenbanken und andere Ressourcen, Suchstrategien, Treffer, Selektionsprozess	Ausgeschlossene Studien mit Begründung
SOP 3	Ergebniszusammenfassung der Qualitätsbewertung	Detaillierte Dokumentation der Qualitätsbewertung
SOP 4	-	Datenextraktionsformulare
SOP 5	tabellarische Ergebnispräsentation	-
SOP 6	Kennzeichnung der Ergebnisse nach der Qualität der zugrunde liegenden Evidenz	-

Diskussion: Im Anschluss an die Darstellung der Entwicklung der SOPs wurden Erfahrungen mit ihrem Einsatz im Rahmen der exemplarischen Bewertung der Ultraschallverfahren beschrieben und diskutiert. Insbesondere sollte festgehalten werden, dass die Evaluation der Validität der hier vorgeschlagenen Methodik im Vergleich zur Standardmethodik der German Scientific Working Group noch aussteht. Idealerweise sollte dieser Abgleich anhand mehrerer Themen vorgenommen werden.

Während der Arbeit mit den SOPs sind vor allem zwei Problembereiche offensichtlich geworden:

Das in SOP 1 vorgeschlagene Abstimmungsverfahren (modifiziertes, email-basiertes Delphi-Verfahren) stellte sich als sehr zeitaufwändig heraus, vermutlich weil die von einzelnen Projektbeiratsmitgliedern wahrgenommene Verantwortung für zügige Themenbearbeitung doch relativ gering war. Diese Einstellung führte zu verzögerten und unvollständigen Rückmeldungen. Wir schließen daraus, dass der Projektbeirat zahlenmäßig auf höchstens vier Mitglieder beschränkt werden sollte. Diese sind verstärkt auf ihre Mitverantwortung für das Gesamtvorhaben hinzuweisen.

Das für SOP 3 entwickelte Instrument zur Bewertung der Qualität von prognostischen Studien zeigte in der Reliabilitätstestung (drei Studien, beurteilt durch sechs Bewerter, Übereinstimmung bei 65%) keine zufriedenstellenden Resultate. Wir nehmen an, dass zu wenig Informationen zum Krankheitsbild und der Technologie zu Unsicherheiten bei der Bewertung führten. Dennoch sollten Instrument und Instruktionen vor einem erneuten Einsatz überprüft werden.

Schlussfolgerungen: Die Erstellung eines systematischen und transparenten Kurz-HTA unter Einsatz der oben skizzierten SOPs sollte innerhalb von drei bis vier Monaten möglich sein. Die Evaluation der Methode gegen einen Goldstandard steht noch aus.

B.2 Exemplarisches „Kurz-HTA“

M Ekkernkamp, D Lühmann, H Raspe

Hintergrund: Die am weitesten verbreitete Definition der Osteoporose geht auf eine Konsensuskonferenz von 1991 zurück: "Eine Erkrankung, die durch niedrige Knochenmasse und eine Beeinträchtigung der Mikroarchitektur des Knochengewebes gekennzeichnet ist und so zu einer erhöhten Knochenbrüchigkeit und daraus folgend zu einem erhöhten Risiko führt, Frakturen zu erleiden". Bisher war die osteodensitometrische Messung der Knochenmineraldichte (KMD) mit Röntgenstrahlen das einzige nicht-invasive Verfahren zur Abschätzung der Knochenbrüchigkeit und damit des intrinsischen Frakturrisikos. Vor dem Hintergrund fehlender ionisierender Strahlung und vergleichsweise geringerer Kosten haben in den letzten Jahren unterschiedliche quantitative Ultraschall-Verfahren (QUS) zunehmende Verbreitung bei der Abschätzung des Risikos für „osteoporotische“ Frakturen gefunden. Sie basieren auf der parametrischen Analyse der Wechselwirkung von mechanischen Schallwellen beim Durchgang durch den Knochen. Zwei verschiedene Ultraschallbasismessgrößen werden hierbei untersucht: die Schallgeschwindigkeit (ultrasound transmission velocity = UTV oder speed of sound = SOS) und die Breitbandultraschallabschwächung (broadband ultrasound attenuation = BUA). BUA und SOS werden nicht nur durch die Knochenmineraldichte (KMD), sondern wahrscheinlich auch durch Knochenstruktur und Knochenzusammensetzung beeinflusst (Barkmann et al. 1999). Vor dem Hintergrund der schnellen und zunehmenden Verbreitung dieser noch jungen Technologie in der Osteoporosedagnostik sollen deren Stellenwert und Wertigkeit in diesem HTA-Schnellverfahren untersucht werden. Dabei liegt der Schwerpunkt auf der Zusammenfassung der Ergebnisse der Frakturrisikoanalyse.

Methodik: Die präzisen Fragestellungen für die Verfahrensbewertung wurden in einem modifizierten Delphi-Verfahren festgelegt. Die hierarchische Literaturrecherche umfasste zwei Schritte: nach HTA-Berichten und systematischen Reviews wurde in den Datenbanken MEDLINE, Health Star, der Cochrane Library und der ISTAHC-Datenbank recherchiert. In einem zweiten Schritt wurde in den Jahrgängen 1998 bis 2/2001 der gleichen Datenbanken nach Publikationen recherchiert, die nach Redaktionsschluss für den einzigen gefundenen HTA-Bericht von 1998 erschienen sind. Die erhaltenen „Treffer“ wurden nach folgenden Relevanzkriterien weiter selektiert:

Relevanz	Kriterien
zentral	Ultraschallmessung, Frakturrisiko, prospektive Studie
peripher	Ultraschall und DXA, Querschnittsstudie, Ultraschall, populationsbezogene Stichprobe: Normwerte, Frakturpatienten im Vergleich zu Personen ohne Frakturen
Ausschluss	Sekundäre Osteoporosen; in vitro Studien, Tierstudien, thematisch abweichend, Briefe, Editorials, Fallberichte; Sprache nicht Englisch, Deutsch, Französisch, Spanisch; eindeutig unsystematische Übersichtsartikel
unklar	anhand der Angaben im Abstract nicht eindeutig zuzuordnen

Die Dokumentation der methodischen Qualität wurde anhand neu entwickelter Stufenpläne/Checklisten für systematische Übersichtsarbeiten und Prognosestudien vorgenommen. Die Informationssynthese erfolgte qualitativ, im Text und tabellarisch.

Ergebnisse: Im modifizierten Delphi-Verfahren wurden folgende präzise Fragestellungen als vordringlich zu beantworten identifiziert (geordnet nach Priorität):

1. Wie valide lässt sich das Frakturrisiko mithilfe der quantitativen Ultraschall-Verfahren im Vergleich zum DXA-Verfahren abschätzen?
2. Welches der unterschiedlichen QUS-Verfahren ist hierzu am besten geeignet?
3. Welche qualitätssichernden Maßnahmen sind dann vor einem Einsatz in der ärztlichen Praxis zu realisieren?
4. Welchen weiteren Forschungsbedarf zur Klärung des Stellenwertes der Ultraschalldiagnostik im Rahmen der Frakturprävention gibt es?
5. Wie wären die unter 1 – 3 ermittelten Ergebnisse zur Patienteninformation und zum Verbraucherschutz zu nutzen?

Zur Beantwortung der Fragen konnte die Literaturrecherche / -selektion einen HTA-Bericht und fünf prospektive Einzelstudien mit zentraler Relevanz auffinden. Die Arbeiten entsprachen den per Checkliste abgefragten Qualitätskriterien.

Der 1998 publizierte kanadische HTA-Bericht von Homik und Hailey, dem eine systematische Literaturübersicht zugrunde liegt, zieht zur Klärung der Frakturvorhersagefähigkeit die Daten aus acht prospektiven Studien heran. Sie kommen zu der Feststellung, dass die meisten Studien eine gute Korrelation von niedrigen Ultraschallmesswerten zur Frakturprävalenz der Studienpopulation zeigten. Das relative Frakturrisiko (RR) verbunden mit niedrigen Ultraschallmesswerten stellt sich ähnlich den korrespondierenden Daten einer DXA-Messung an der Wirbelsäule und Hüfte dar und lag in multivariaten Analysen zwischen 1,1 und 1,6. Eine Studie fand ein relatives Risiko von 2, hier wurde allerdings die Wirbelkörperdeformität als Zielgröße betrachtet und nicht das klinische Frakturereignis. Die Autoren des HTA-Berichtes kamen zu der Schlussfolgerung, dass das quantitative Ultraschallverfahren (QUS) im Fersenbeinbereich eine vielversprechende diagnostische Technologie darzustellen scheint, dessen Rolle in der Osteoporosedagnostik jedoch noch nicht eindeutig definiert ist. Als alleiniger Parameter zur Abschätzung des Frakturrisikos ist die Ultraschallmessung nicht geeignet. Die kanadischen Autoren formulieren weiteren Forschungsbedarf hinsichtlich Langzeit-Präzision, Frakturrisikoabschätzung und potentielle Kosten-Effektivität der Verfahren zur Klärung des Stellenwertes der QUS-Verfahren in der alltäglichen Gesundheitsversorgung.

Von den weiteren fünf identifizierten prospektiven Studien erfüllten vier alle acht in der Validitätsprüfung abgefragten Qualitätskriterien. Ansonsten war unter ihnen noch eine erhebliche epidemiologische, klinische und technische Heterogenität festzustellen. Die Studienpopulationen z. B. waren so heterogen, dass ein direkter Vergleich in nur sehr eingeschränktem Umfang möglich war. Auch bei der Expositionsmessung, die zwar in allen fünf Studien transparent dokumentiert wurde und intern den An-

sprüchen an Reproduzierbarkeit und Validität genügte, war erhebliche Heterogenität festzustellen. Vier Verfahren unterschiedlicher Gerätehersteller, die sich in Messort (Kalkaneus, Radius, Patella), Schalleinleitung (Wasser, Gel) und Messgrößen (SOS, BUA, UTV) unterschieden, liefern nicht direkt vergleichbare Werte. Eines der eingesetzten Verfahren ist bereits technisch überholt und obsolet. In vier Studien war die Hüftfraktur primäre Zielgröße, in zwei Studien sogar einzige Zielgröße. Lediglich eine populationsbezogene Untersuchung nimmt keine Auswertung von Hüftfrakturereignissen vor, sondern analysiert die Daten nach Wirbel- und nicht-spinalen Frakturen. Auch die berichteten Effektgrößen unterschieden sich in den Studien erheblich. In einer Publikation werden Angaben zur Sensitivität des QUS-Verfahrens zur Erkennung von Frakturgefährdung gemacht. Da es sich um prospektive Daten handelt, können diese Größen als direkte Aussage zum prognostischen Wert interpretiert werden. Vier Studien gaben adjustierte relative Risiken an. Dabei wurden in einer Studie nur das Alter, einer weiteren Studie zusätzlich Gewicht und Geschlecht und in der letzten Studie Alter, Gewicht, Geschlecht, Körpergröße und laufende Osteoporosetherapie berücksichtigt.

Diskussion der Forschungsfragen:

1. Eine prognostische Gleichwertigkeit der Ultrasonometrie gegenüber DXA wurde belegt; QUS und DXA ermitteln, jeweils für sich betrachtet, vergleichbare relative Risiken für Frakturen innerhalb eines Beobachtungszeitraums von $\frac{1}{2}$ bis $5 \frac{1}{2}$ Jahren. Die Meta-Analyse von Marshall et al. (1996) berichtet für Knochendichtewerte pro Erniedrigung um je 1 Standardabweichung ein relatives Risiko von 1.6 für alle Frakturen und alle Messorte. Der in den fünf neueren prospektiven Studien berichtete Bereich der relativen Risiken (bezogen auf eine Differenz von - 1 SD) liegt für die BUA-Messungen in der gleichen Größenordnung (von 1.17 bis 2.3). Knochendichte und Ultraschallparameter scheinen von einander unabhängige Risiken zu bezeichnen (Homik & Hailey 1998). Die Kombination von QUS und DXA bei multivariater Betrachtung führt möglicherweise zu einer verbesserten Frakturprädiktion, die uns vorliegende Datenbasis zu dieser Frage ist allerdings begrenzt. Empfehlungen, welche Verfahren und Messorte zur Prädiktion welcher Frakturtypen herangezogen werden könnten, lassen sich aus diesen Daten allerdings noch nicht ableiten, hier besteht sicherlich weiterer Forschungsbedarf.
2. Die Eignung von Ultraschallverfahren ist abhängig von der Zuverlässigkeit (Richtigkeit und Präzision) der Messungen und - in Abhängigkeit vom Messergebnis - der Höhe und Sicherheit des geschätzten relativen Risikos für Frakturen. Aus praktischer Sicht sind schließlich die Verfügbarkeit und Verbreitung der Technik zu beachten. Berücksichtigt man Reliabilität, prognostischen Wert und Verbreitung gleichzeitig, dann spricht vieles für die transversale Messung von BUA/SOS am Kalkaneus.
3. Eine wichtige Voraussetzung für die Einführung einer quantitativen Ultraschallmethode zur Knochenuntersuchung in die klinische Praxis in Deutschland sollte das Vorhandensein eines genügend großen und repräsentativen Referenzkollektivs gesunder Frauen und Männer aus unserer europäischen Region sein. Dabei ist zu be-

achten, dass Referenzwerte für alle einzuführenden Systeme und für jede Messlokalisierung getrennt zu erheben sind. Weiterhin muss die für die DXA-Methode etablierte WHO-Krankheitsdefinition ("Osteoporose" bei Messwerten von $< - 2,5$ SD unter dem Mittelwert junger gesunder Frauen) für das einzuführende QUS-Verfahren erneut validiert werden.

Größte Sorgfalt muss der Schulung der die Messung Durchführenden und Beurteilenden gewidmet werden, da diese durch falsche Lagerung des Patienten, Bewegungen des Patienten während der Messung oder sonstige die Messung beeinträchtigende äußere Umstände wie Raumtemperatur, Kleidung und anderes die Messergebnisse verfälschen können. Die Schulung sollte unter standardisierten Bedingungen an Kompetenzzentren durchgeführt werden und dem Erwerb eines Befähigungsnachweises abschließen.

4. Angesichts der vielfältigen Unklarheiten und Unsicherheiten im Bereich der QUS-Verfahren scheint eine Fortführung und Intensivierung der medizinischen und physikalischen Grundlagen- und krankheitsbezogenen Forschung erforderlich. Dies würde auch helfen, das diagnostische Potential von QUS zu klären. Von großer Bedeutung sind anschließende Interventionsstudien. Bislang liegen nur valide Daten zur Vorhersagbarkeit von Frakturen bei Personen vor, die bislang **nicht** mit knochenwirksamen Substanzen behandelt wurden. Prinzipiell ist festzuhalten, dass eine diagnostische Maßnahme nur dann sinnvoll ist, wenn zumindest die Option einer aus dem Messergebnis sich ableitenden – effektiven und effizienten – Therapie besteht.

5. Dringend gebraucht werden ausgewogene Verbraucherinformationen, die dem potentiellen Nutzer des Verfahrens eine realistische Einschätzung der Aussagekraft der Messergebnisse ermöglichen. Vor allem sollte auch den Verbrauchern gegenüber klar gestellt werden, dass sich das Verfahren noch in der Entwicklung befindet, dass die therapeutischen Konsequenzen erniedrigter Befunde unklar sind und dass derzeit Verlaufskontrollen kaum möglich sind.

Schlussfolgerungen:

- Solange der Gegenstand der verschiedenen QUS-Messungen nicht klar ist, kann die Frage nach dem Beitrag des Verfahrens zur Osteoporosediagnose nicht beantwortet werden. Bei DXA-Messungen hat sich international der Gebrauch von T- und Z-Werten eingebürgert, um zu kategorialen Urteilen (z. B. „Osteopenie“, „Osteoporose“ nach WHO) zu kommen. Davon scheinen die QUS-Messungen noch weit entfernt. Es besteht dringender Bedarf zu weiterer medizinischer und physikalischer Forschung.
- Es besteht zum jetzigen Zeitpunkt keine klare Indikation, Ultraschalluntersuchungstechniken in der Osteoporose-Diagnostik einzusetzen. Hierüber müssen Verbraucher und Patienten aufgeklärt werden.
- Prinzipiell scheint ein zukünftiger Einsatz von QUS-Verfahren für die Frakturrisikoabschätzung möglich. Ihr Stellenwert für die Risikoerkennung und klinische Prävention ist derzeit aber noch unsicher. Der Gebrauch sollte daher an wissenschaftliche Fragestellungen und vor allem auch an eine konsequente Qualitäts-

sicherung gebunden sein. Über Abweichungen hiervon sind diejenigen zu informieren, die das Verfahren, etwa als IGEL-Leistung, in Anspruch nehmen wollen.

- Die Datenlage für einen Einsatz der QUS-Verfahren in der Verlaufskontrolle scheint begrenzt und kontrovers. Sichere Aussagen sind zur Zeit nicht möglich, auch weil die therapeutische Beeinflussung und Veränderungssensitivität der QUS-Parameter zu wenig geklärt sind. Einschränkend ist zu betonen, dass dieser Teilaspekt nicht Gegenstand unseres HTA-Schnellverfahrens war.

I C Hauptdokument

I C.1 Zielsetzung des Berichtes

Ziel des hier berichteten Projektes war die Entwicklung von standardisierten Vorgaben für die Durchführung von Kurz-HTAs und die Erstellung von Kurz-HTA-Berichten unter strikter Beibehaltung des systematischen Ansatzes und der absoluten Transparenz.

I C.2 Hintergrund

I C.2.1 Was ist Health Technology Assessment (HTA)?

Unter Health Technology Assessment (HTA) wird die umfassende Bewertung (und ggf. Evaluation) neuer oder auf dem Markt befindlicher Technologien (Arzneimittel, Medizinprodukte, Dienstleistungen, Verfahren sowie Organisationssysteme) der gesundheitlichen Versorgung hinsichtlich ihrer medizinischen, physikalischen, biologischen, sozialen und finanziellen Wirkungen im Rahmen einer strukturierten Analyse verstanden (Bitzer et al. 1998; Perleth 2000). Seine Aufgabe ist die Bereitstellung und Verteilung von wissenschaftlich fundierten Informationen für Kommunikations- und Entscheidungsprozesse auf verschiedenen Ebenen und bei unterschiedlichen Institutionen des Medizin- und Gesundheitssystems. Hierzu zählen beispielsweise Investitionsentscheidungen, Kostenübernahmeentscheidungen und die Gestaltung von Leistungskatalogen der Kostenträger (Banta & Luce 1993).

HTA gehört damit zu den Konzepten einer "Evidenz-basierten Gesundheitsversorgung", die für jegliche Entscheidung eine systematisch erarbeitete und transparent dokumentierte Entscheidungsgrundlage fordert. HTA und EbCM (Evidence-based Clinical Medicine) bedienen sich identischer Erkenntnismittel, haben eine übereinstimmende Qualitätsskala für Evidenzen, stimmen teilweise hinsichtlich der Elemente oder Kriterien überein, unterscheiden sich aber grundsätzlich in der Zielorientierung ihrer Empfehlungen. Bei EbCM geht es um eine Verwertung von wissenschaftlicher Evidenz in der klinischen Versorgung von Individuen oder umschriebenen Gruppen (Leitlinien), bei HTA um die Abgabe von Empfehlungen auf der Systemebene (Forschung und Entwicklung, Regulierung von Arzneimitteln und Medizinprodukten, Steuerung der Anzahl und Standorte von medizinischen Leistungen, Kostenübernahme von Leistungen, Qualitätssicherung, Aus- und Weiterbildung der Anbieter, Konsumenteninformation u. a.).

Typischerweise geht es bei der Technologiebewertung um folgende inhaltliche Aspekte (Bitzer et al. 1998):

- Sicherheit (Phase I)
- Experimentelle Wirksamkeit (Phase II; efficacy)
- Wirksamkeit unter Alltagsbedingungen (Phase III; effectiveness)
- Vergleichende Bewertung der Wirksamkeit (Phase IV; comparative effectiveness)
- Kosten-Wirksamkeit
- Soziale, rechtliche und ethische Implikationen

HTA-Berichte beziehen sich nicht zwangsläufig immer auf alle Elemente von HTA, sondern enthalten meistens nur Teil-Bewertungen, könnten also z. B. Kosten-Wirksamkeits-Untersuchungen auslassen (Hart 2001).

HTA hat sich in vielen Gesundheitssystemen als wichtiges Instrument zur Unterstützung von gesundheitspolitischen Entscheidungen etabliert, welches in unterschiedlichen Steuerungszusammenhängen und zur Politikberatung einsetzbar ist. Übergeordnete Ziele sind dabei:

- die Zulassung neuer Technologien,
- die Anwendung zugelassener Technologien und
- die Koordinierung von Forschungsaktivitäten

(Perleth / Busse / Schwartz 1999).

Ausschlaggebend für die Erstellung von HTA-Berichten ist oftmals die zunehmende Nutzung von neuen medizinischen Technologien, ohne dass Genauerer über den Nutzen des Verfahrens bekannt ist oder das Auftauchen neuerer Indikationen für bereits in der Routineversorgung etablierte Verfahren (Perleth & Wild 2001). In diesen Situationen kann ein HTA-Bericht auf der Basis einer systematischen Literaturübersicht wichtige Informationen liefern und ist möglicherweise auch eine Grundlage für die Entwicklung von Leitlinien, Patienteninformationen und für Entscheidungen im klinischen Alltag.

Internationale Erfahrungen zeigen, dass HTA-Berichte dann den größten Einfluss ausüben, wenn sie einen konkreten Bedarf befriedigen, also angefordert wurden oder auf aktuellen Entscheidungsbedarf zugeschnitten sind. In verschiedenen Ländern haben sich Mechanismen etabliert, die ständig neue Entwicklungen in der Medizin und ihre potenzielle Relevanz für das Gesundheitssystem beobachten (sog. "horizon scanning") (Hailey et al. 2001).

Für bereits eingeführte Technologien hebt Perleth (2001) als mögliche Konsequenz der Ergebnisse von HTA-Berichten zwei Extreme hervor: keine Änderung (des Technologiegebrauchs) oder komplette Blockade einer Technologie. Zu den zwischen diesen Extremen liegenden Möglichkeiten gehören die Beschleunigung der (Weiter)entwicklung einer Technologie, Änderung der Indikationen für die Anwendung einer Technologie, Nutzung der Informationen für die strategische Planung, stufenweise Einführung oder zeitlich begrenztes Aussetzen der weiteren Diffusion bei gleichzeitiger weiterer Evaluation der Technologie, Initiierung weiterer Forschung hinsichtlich Risiken oder Wirksamkeit, Monitoring der Nutzung der Technologie.

In Deutschland ist mit Wirkung ab dem 1.1.2000 durch Artikel 19 GKV-Gesundheitsreformgesetz 2000 eine gesetzliche Grundlage für ein Informationssystem zur Bewertung medizinischer Technologien geschaffen worden, dessen Ziel es ist, „ein datenbankgestütztes Informationssystem für die Bewertung der Wirksamkeit oder der Effektivität sowie der Kosten medizinischer Verfahren und Technologien“

beim DIMDI zu errichten und zu betreiben, das „den Zugang zu den relevanten Datenbanken“ erschließt und „Studien und sonstige Materialien zum Stand der nationalen und internationalen wissenschaftlichen Erkenntnisse im Bereich der Technologiebewertung in der Medizin“ erfasst.

Diese Gesetzesänderung bildet den vorläufigen Höhepunkt der Entwicklung von HTA in Deutschland, die 1995 mit einer Projektförderung des Bundesministeriums für Gesundheit (Projekt „Aufbau / Weiterentwicklung einer Datenbasis: Evaluation medizinischer Verfahren und Technologien in der Bundesrepublik Deutschland“) begann. Ziel dieses Projektes war es, die Methodik der Erstellung von HTA-Berichten in Deutschland zu etablieren, um eine beim DIMDI angesiedelte Informationsdatenbank aufzubauen. Die Projektzentrale befand sich zunächst am Institut für Sozialmedizin, Epidemiologie und Gesundheitssystemforschung (ISEG) in Hannover, später an der Medizinischen Hochschule Hannover, die kooperierenden Arbeitsgruppen waren auf unterschiedliche deutsche Universitäten und Forschungsinstitute verteilt (Lübeck, Wuppertal, Ulm, München, Hannover, Bochum, Freiburg u. a.). Die gesamte Projektgruppe nannte sich, auch um international identifizierbar zu sein, "German Scientific Working Group for Health Technology Assessment". Die Projektarbeit umfasste folgende Aktivitäten, deren Ergebnisse als Buchreihe (Health Technology Assessment. Schriftenreihe des Deutschen Institutes für medizinische Dokumentation und Information im Auftrag des Bundesministeriums für Gesundheit) im Nomos-Verlag erschienen sind:

- eine Bestandsaufnahme von internationalen HTA-Aktivitäten mit Entwicklung einer Konzeption zur Implementation von HTA in Deutschland;
- die Sammlung, Bewertung und Anpassung von ausländischen Verfahrensbewertungen zur Verwendung im Kontext des bundesdeutschen Gesundheitssystems;
- die genuine Bewertung von medizinischen Technologien und Prozeduren, wobei im Rahmen der verfügbaren Literatur neben der medizinischen Wirksamkeit auch die Kosten-Wirksamkeit berücksichtigt wurde;
- die Überprüfung der Einzelberichte auf Aktualität. Falls eine Überarbeitung notwendig erscheint, wird diese in Absprache mit dem Auftraggeber und unter Berücksichtigung des jeweils anfallenden Aufwandes sowie der zur Bearbeitung anstehenden neuen Themen durchgeführt;
- die Standardisierung der Methoden zur Erstellung von HTA-Berichten.

Die im Rahmen der "German Scientific Working Group" gegründete Arbeitsgruppe „Methoden der Technologiebewertung“ versuchte, für die verschiedenen Aspekte der Technologiebewertung ein einheitliches methodisches Vorgehen zu definieren bzw. internationale Standards zu übernehmen und ggf. zu adaptieren. Diese Vorgaben wurden als sogenanntes "Toolkit" arbeitsgruppenintern aufgezeichnet und neuen Projektmitarbeitern als Arbeitsanleitung an die Hand gegeben.

Das von der Deutschen Arbeitsgruppe HTA entwickelte Konzept schlägt folgende Gliederung für eine Verfahrensbewertung vor (Bitzer et al. 1998):

- Definition des zu klärenden Problems auf der Seite des Adressaten / Auftraggebers
- Definition der Forschungsfrage
- Beschreibung der Ausgangssituation
- Technische Charakteristika des Verfahrens (z. B. Präzision)
- Daten zur experimentellen Wirksamkeit (efficacy)
- Daten zur Wirksamkeit unter Alltagsbedingungen (effectiveness)
- Im Verhältnis (Vergleich) zu anderen Verfahren
- Impact der Technologie auf die Organisation der gesundheitlichen Versorgung
- Ökonomische Evaluation (efficiency)
- Ethische, rechtliche, soziale und psychologische Implikationen
- Schlussfolgerungen, Optionen, Empfehlungen.

Internationale Berichte, aber auch die Erfahrungen aus mehr als 20 bisher in Deutschland bearbeiteten und publizierten Themen haben gezeigt, dass pro Thema (einschließlich Reviewverfahren) eine Bearbeitungszeit bis zum Erscheinen der Publikation von mindestens 12, oft auch 18 Monaten zu veranschlagen ist.

Vor diesem Hintergrund scheint es sinnvoll, ein Konzept für sogenannte „schnelle Technologiebewertungen“ (rapid assessments) zu erarbeiten. Ziel ist es, innerhalb eines kurzen Zeitraumes von 4 (3 - 6) Monaten einen HTA-Bericht zu erstellen, der evidenzbasierte Informationen zur Entscheidungsunterstützung bei dringenden Fragestellungen liefern kann.

Im Folgenden sollen zunächst der Begriff „rapid assessment“ geklärt und die Konzepte für ausführliche und schnelle HTAs anhand einer Komponentenanalyse gegenüber gestellt werden.

I C.2.2 Was ist ein „rapid assessment“?

Kernstück jeglicher evidenzbasierter Entscheidungsgrundlage, und damit auch von HTA-Verfahren, ist die umfassende und systematische Zusammenfassung der Evidenzlage in einer Literaturübersicht. Hierfür wird der systematische Review (ggf. mit Meta-Analyse) allgemein als Goldstandard angesehen. Wegen des großen Zeit- und Kostenaufwandes ist diese Methode in der Regel aber für dringende Entscheidungen

nicht praktikabel. Vor allem begrenzte Ressourcen für die gesundheitliche Versorgung und immer neue, auf den Markt drängende therapeutische oder diagnostische Technologien werden voraussichtlich einen zunehmenden Bedarf an zeitnahen Verfahrensbewertungen bedingen.

Hierbei wird es von wesentlicher Bedeutung sein, vor der Etablierung der neuen Maßnahme einen Evidenzbericht zur Verfügung zu stellen, um entscheiden zu können, ob die Verbreitung der Maßnahme gefördert oder gebremst werden sollte (Ziegler et al. 2001). Neben Innovationen können aber auch neue Indikationen, Veränderungen von Kostenstrukturen, geändertes Nachfrageverhalten, demographische Veränderungen und ähnliche Faktoren einen Problemdruck schaffen, der eine zügige Entscheidung für oder gegen den Einsatz eines Verfahrens erforderlich macht.

Während die Erstellung von umfangreichen HTA-Berichten mindestens 12, eher 18 Monate Zeit erfordert, müssen Entscheidungen aber oft innerhalb von Wochen bis wenigen Monaten getroffen werden. Um relevante Informationen auch für rasche Entscheidungsfindungen zur Verfügung stellen zu können, wurden in den letzten Jahren, vorwiegend innerhalb von bestehenden HTA-Programmen, zusätzliche sogenannte „rapid assessments“, Schnellverfahren, entwickelt. Derzeit lassen sich allerdings noch keine einheitlichen Definitionen von Vorgehensweise und Zeitspannen ableiten. Nach Perleth et al. (2001) könnte man z. Zt. rapid assessments wie folgt beschreiben:

Schnelle HTAs bearbeiten eine fokussierte Fragestellung, sind auf maximal 6 Monate angelegt und greifen sowohl auf Primär- als auch auf Sekundärstudien zurück.

Um HTA als transparentes und valides Politikberatungsinstrument auch für dringliche Entscheidungen nutzbar zu machen, sollte nach anderen Autoren – insbesondere aus dem englischen Sprachraum - zufolge ein Schnellverfahren:

- in der Regel nach 3 - 4 Monaten Bearbeitungszeit abgeschlossen sein,
- auf den Informationsbedarf der Nutzer zugeschnitten sein,
- zur Sicherstellung von absoluter Transparenz die Themenbearbeitung in systematischen und standardisierten Arbeitsschritten vornehmen.

International haben sich in einigen HTA-Einrichtungen Schnellverfahren für die Verfahrensbewertung etabliert, die jedoch sehr unterschiedliche Ansätze haben. Vor allem die Erfahrungen in England zeigen, dass für ein erfolgreiches Programm stabile Strukturen in Verbindung mit Routineprozessen in der Durchführung von schnellen HTAs und insbesondere der Verfügbarkeit eines erfahrenen Reviewteams notwendig sind. Als Beleg hierfür mögen die seit Jahren erfolgreich arbeitenden Development and Evaluation Committees (DECs) in Birmingham, Southampton und Sheffield dienen. Dabei handelt es sich um Arbeitsgruppen, die auf der Basis von methodischen Handbüchern weitgehend standardisiert schnelle HTAs erstellen (Stevens et al. 1995).

In Tabelle 1 sind die Charakteristika von einigen bekannten schnellen HTA-Programmen zusammengestellt.

Tabelle 1: Charakteristika schneller HTA-Programme (nach Perleth et al. 2001)

Programm/ Institution	Dauer	Methoden	Perspektive/ Zielgrößen	Anzahl
DES Birmingham	3 – 6 Monate	Auswertung von Primärstudien	Wirksamkeit, Lebensqualität	k. A.
DEC Southampton	k. A.	Primärstudien, Reviews, Metaanalysen	Wirksamkeit, Lebensqualität	Bisher ca. 200
Trent Sheffield	k. A.	k. A.	k. A.	Ca. 8/Jahr seit 1996
ARIF Birmingham	Bis 4 Wochen	Systematische Reviews	k. A.	Ca. 70/Jahr
NCCHTA Southampton	2 – 3 Tage	Primärstudien, systematische Reviews aus Medline und Coch- rane	k. A.	Bisher ca. 30
CAHTA Spanien	k. A.	Primärstudien	k. A.	k. A.
CCOHTA Kanada	k. A.	Überwiegend für Arz- neimittel; Primärstudien	k. A.	Ca. 3-4/Jahr seit 1997
AHFMR Kanada	2 – 3 Monate	Andere HTAs; Primärstudien	i.d.R. keine Kos- ten	Ca. 5/Jahr seit 1996
ECRI USA Windows HTA	3 – 6 Monate	Systematischer Review; Metaanalysen,einfache statistische Verfahren	Wirksamkeit, Kosten, Regula- tions-, Diffusi- onsstatus	Ca. 15-20/Jahr Seit 1998
Target database reports	6 – 8 Wochen	Zusammenstellung aus Auszügen einer HTA- Daten-Bank	Schwerpunkt auf klinischen Stu- dien	120 Reports bis Ende 1999
Hotline responses	< 5 Tage	Abstracts aus Literatur- recherche, ohne Aus- wertung	k. A.	k. A.
Neuseeland Level 1	k. A.	Primärstudien, Reviews, Informationszusammen- fassung, internes Re- view	k. A.	Insgesamt 11 Re- ports seit 1997;k. A. zur Zuordnung zu Levels
Level 2	k. A.	L1 + internes Critical Appraisal + Experten- konsultation Schlussfolgerung: Emp- fehlungen mit Angabe der Evidenzbasis	k. A.	

Programm/ Institution	Dauer	Methoden	Perspektive/ Zielgrößen	Anzahl
Level 3	k. A.	L1 + 2 + externes Review der Evidenz durch Expertenpanel; Schlussfolgerung: Empfehlungen mit Angabe der Evidenzbasis	k. A.	
Level 4	k. A.	Wie L3 + extensive Recherche nach grauer Literatur und laufenden Studien	k. A.	

In Deutschland sind nachfolgend aufgelistete unterschiedliche Anwendungsoptionen von rapid assessments bzw. Kurz-HTAs bzw. Kurz-Stellungnahmen zur Evidenzlage denkbar, wobei diese nicht als kostengünstige Minimalvariante eines ausführlichen HTA-Berichtes verstanden werden dürfen:

- Vorläufige Definition von Leistungskatalogen
- Frühzeitige Identifikation von kostenintensiven Innovationen ("Horizon Scanning")
- Priorisierung von Beratungsthemen (Festlegung der Reihenfolge, Umfang des Beratungsthemas, Abwarten relevanter Forschungsergebnisse)
- Aktualisierung eines ausführlichen HTA-Berichtes

Prinzipiell sind zwei Lösungswege zur Durchführung eines rapid assessment vorstellbar:

- Durch vermehrten Ressourcenaufwand werden ausführliche HTA-Berichte schneller erstellt.
- Durch Beschränkung der Evaluationsbreite und -tiefe werden fokussierte HTA-Berichte mit geringerem Zeitaufwand erstellt.

Ein Konzept für ein methodisch einwandfreies, transparentes und nutzer- bzw. bedarfsorientiertes Kurz-Assessment im Kontext des bundesdeutschen Gesundheitssystems stellt das skizzierte „Bedarfsadaptierte Schnellverfahren zur Erstellung von Health Technology Assessments“ dar.

I C. 2.3 Notwendigkeit und Konzept eines Schnellverfahrens

Die Durchführung von Health Technology Assessments in der deutschen HTA-Arbeitsgruppe erfolgte bisher nach einem methodischen Vorgehen wie im Endbericht „Bestandsaufnahme, Bewertung und Vorbereitung der Implementation einer Daten-

sammlung ‚Evaluation medizinischer Verfahren und Technologien‘ in der Bundesrepublik“ beschrieben (Bitzer et al. 1998).

Als Qualitätsindikatoren für HTAs stehen v. a. systematische Methodik, Vollständigkeit und absolute Transparenz (Perleth et al. 2000). Ein HTA-Bericht umfasst danach die Beschreibung des entscheidungsrelevanten Kontext, systematisch zusammengestellte Informationen zur Zielkondition und zum Verfahren und seinem derzeitigen Einsatz, die systematische Aufarbeitung der gesamten verfügbaren wissenschaftlichen Literatur zur Effektivität unter Studien- und Alltagsbedingungen sowie zu unerwünschten Wirkungen. Hierdurch wird der hohe Zeitbedarf für die Fertigstellung von Bewertungen determiniert. Um eine Themenbearbeitung in einem überschaubaren Zeitraum von weniger als sechs Monaten zu ermöglichen und gleichzeitig ein methodisch einwandfreies, transparentes und nutzer- bzw. bedarfsorientiertes Kurz-Assessment durchzuführen, sollten Standardprozeduren festgeschrieben werden. Die in der Folge vorgestellte Komponentenanalyse kann diesen Prozess strukturieren und erleichtern.

Die Konzeption des bedarfsadaptierten Kurz-Assessments beruht auf einer zweidimensionalen Analyse des HTA-Prozesses (siehe Tabelle 2). Die erste Dimension (horizontal) beschreibt inhaltliche Bestandteile des Verfahrens (Vollständigkeitsanspruch). Die zweite Dimension (vertikal) nimmt für jede inhaltliche Komponente eine Hierarchisierung (Vertiefungsstufen) möglicher methodischer Arbeitsschritte nach Aufwand und Komplexität vor.

Tabelle 2 zeigt zusammenfassend Unterschiede zwischen Standard-HTA (dunkelgrau), Maximalvariante (weiß) und Kurz-HTA (hellgrau), sowohl was die Breite der Themenbearbeitung als auch den methodischen Aufwand für die einzelnen Arbeitsschritte betrifft.

Tabelle 2: Komponenten eines HTA-Verfahrens

Hintergrund				Sicherheit, Wirksamkeit				fakultative Komponenten*		Schlussfolgerungen	Anhang	Peer Review
Technologie	Zielkondition	Epidemiologie	Technologie-nutzung	Literatur-recherche	Critical appraisal	Daten-extraktion	Informations-synthese	ökonomische Analyse ***	psycho-soziale und ethische Implikationen			
Funktions-prinzip	Pathophy-siologie, Klinik, Verlauf	aus Litera-tur: Bücher, Übersichten	Expertise (eigene, ausgewähl-te Experten)	HTA-Datenbanken	Definition von inhaltlichen Ein- u. Aus-schlusskrite-rien	Datenex-trakti-onsformu-lar	narrativ, Haupt-ergebnisse	Kosten-aufzählung		narrativ, diskutierend	Literaturver-zeichnis	ohne
Abhängig von der Technologie: technische Informationen (diagnost. Tests: Accuracy; Arzneimittel: Galenik, Pharmako-kinetik usw.)	diagnosti-sche, the-rapeutische Ansätze (Methoden wie oben)	aus Litera-tur: syste-matisch re-cherchiert: Primär-studien	lt. evidenz-basierten Leitlinien	Cochrane Library, Medli-ne nach Re-views (letzte 3-5 Jahre)	methodische Ein- u. Aus-schluss-kriterien: Sortierung nach Evidenz-level in An-passung an die Frage-stellung	2 unabhän-gige Daten-extraktionen	Ergebnistabellen			wie oben + Diskussion von Schlussfolgerungen anderer, Empfehlungs-papiere (Kon-sensusstate-ments, Gui-delines etc.)	Methoden-dokumenta-tion: Re-cherche, Appraisal-Instrumen-te; externe Informati-onsquellen	interner Review**, informell
Standards, Zulassungs-modalitäten	Inhalte wie oben, aber systema-tisch re-cherchiert	Gesund-heits-berichter-stattung	Routineda-ten (Kran-kenkassen etc.)	Kerndaten-banken (Med-line, Healthstar, Embase): Originalarbei-ten (letzte 3-5 Jahre)	Methodische Anforderungen: simple Checkliste (z.B. Jadad-Score)	Berechnung einheitlicher Effektstär-ken	qualitative Ergeb-nis-zusammenfas-sung: a) methodische Limitationen b) Studienpopulati-onen c) Interventionen d) Settings und Modifikations-faktoren e) Outcomes (Art, Robustheit, Be-deutung) f) Lebensqualität			gewichtet nach Evi-denzstärke (z.B. wie AHRQ-Leitlinien)	Dokumenta-tion der gesamten Recherche und Apprai-sal-ergebnisse	Fest-schrei-bung von regel-mäßigen internen Reviews - Aktualisierungs-bedarf

Qualitätskontrollen, Wartungsbedarf		Aufarbeitung von Sekundärdaten (z.B. administrativ, Register)	Instrumente der Marktforschung: Survey, Fokusgruppen	Kerndatenbanken seit 1966	Methodische Anforderungen: komplexe Checkliste (z.B. Hadorn, HTA-Arbeitsgruppe)		Metaanalyse wenn anwendbar, ggf. mit Subgruppenanalysen					externes Peer Review, informell
Praxisrelevante Aspekte: Einbindung in Arbeitsabläufe; Zeit, Zusatzausrüstung, Ausbildungsbedarf			Analyse der treibenden Kräfte	Bibliothekar mit EBM-Erfahrung, z.B. Cochrane Gruppe, DIMDI SOP	übergreifend: 1 vs. mehrere Beurteiler; Interrater-Reliabilität, verblindetes Appraisal		Modellrechnungen: Extrapolation von Effekten auf andere Populationen; ggf. Zusammenführen von Informationen aus unterschiedlichen Quellen					formelles, dokumentiertes internes und externes Reviewverfahren
				Handsuche relevanter Zeitschriften und Kongressbände, Referenzlisten, unveröffentlichte Studien			Entscheidungsanalyse					
		Survey		Einbezug von Routedaten	systematische Entwicklung von Fragestellungsspezifischen Appraisal-Instrumenten inklusive Validierung	Anforderung von Originaldaten	Metaanalyse basierend auf Originaldaten			formalisiertes Verfahren zur Integration von Positionen anderer Interessengruppen	alle Originalquellen, z.B. Links bei Online-Publikationen	
Schnellverfahren	Standardverfahren	bisher nicht vorgesehen										

* abhängig u.a. von der Art der Technologie

** Review durch zwei Mitglieder der German Scientific Working Group Technology Assessment for Health Care

*** im Rahmen eines HTA zur medizinischen Effektivität sind keine weiteren ökonomischen Analysen vorgesehen.

Um Praktikabilität, Verwertbarkeit und Akzeptanz sicherzustellen, haben wir die Entwicklung der Verfahrensschritte in einer kooperativen Arbeitsgruppe aus Methodikexperten, Klinikern und zukünftigen Nutzern des Kurz-HTAs vorgenommen. Vorschläge für methodische Vorgehensweisen werden in Abgleich mit der bisher im Projekt „Evaluation medizinischer Verfahren und Technologien“ verwendeten Methodik ("Toolkit") und der internationalen Literatur diskutiert. Im Kern wird es darauf ankommen, die bisher gebräuchlichen Arbeitsschritte einzuengen und zu verkürzen und gleichzeitig sicherzustellen, dass eine derartige Vorgehensweise zu keiner systematischen Verzerrung der Schlussfolgerungen führt. Die Entwicklung dieser methodischen, in der Folge beschriebenen Arbeitsschritte wird an einem Thema mit gesundheitspolitischer Relevanz aber ohne konkreten Entscheidungsdruck vorgenommen. Es geht um „Die Rolle der quantitativen Ultraschallverfahren zur Ermittlung des Risikos für osteoporotische Frakturen“.

I C.2.4 Beschreibung der Arbeitsabläufe/Vorgehen

In Deutschland wurde das vom Bundesministerium für Gesundheit geförderte Projekt „Informationssystem Gesundheitsökonomische Evaluation“ an der Etablierung einer Methodik für Verfahrensbewertungen (HTA) beteiligt. Methodische Vorgehensweisen für ausführliche HTAs sind teilweise beim DIMDI abrufbar (SOPs Literaturrecherche <http://www.dimdi.de>) bzw. werden teilweise in einem Methodenband der Reihe Health Technology Assessment „Bestandsaufnahme, Bewertung und Vorbereitung der Implementation einer Datensammlung – Evaluation medizinischer Verfahren und Technologien - in der Bundesrepublik Deutschland“ publiziert und sind außerdem im arbeitsgruppeninternen Methodikleitfaden (Toolkit: Informationsmaterial für Verfasser von HTA-Berichten) enthalten. Diese Methodik bildet die Grundlage für die hier vorgeschlagene Vorgehensweise im Rahmen von HTA-Schnellverfahren.

Die Entwicklung der methodischen Standards für HTA-Schnellverfahren erfolgte in Kooperation mit der

- Methodenarbeitsgruppe der „German Scientific Working Group for Technology Assessment in Health Care“
- Potentiellen Nutzern von HTA-Berichten (MDK, MDS, Bundesausschuss Ärzte und Krankenkassen).

Die Methodik zur Erstellung eines Kurz-HTAs lässt sich in sieben zentrale Arbeitsschritte gliedern, für die die methodischen Vorgehensweisen in sogenannten SOPs (Standardized Operating Procedures) zusammengefasst werden sollen. Im Einzelnen handelt es sich dabei um:

- SOP 1: Systematische Festlegung der präzisen Berichtsfragestellung unter Einbindung von Nutzern und Fachgebietsexperten
- SOP 2: Informationsrecherche

- SOP 3: Informationsselektion und –bewertung
- SOP 4: Informationsextraktion
- SOP 5: Informationssynthese
- SOP 6: Generierung von Schlussfolgerungen, Aktualisierungsbedarf
- SOP 7: Dokumentationspflichten

Ausgangspunkt für jedes SOP ist das bisher gebräuchliche methodische Vorgehen der German Scientific Working Group, wie es im oben erwähnten „Toolkit“ von Juni 1999 festgeschrieben wurde, darüber hinaus dienen als Basis für SOP 2 die vom DIMDI erarbeiteten Standards „Datenbasis“ und „Literaturrecherche“. Einige ausländische HTA-Institutionen verfügen bereits über mehrjährige Erfahrungen mit der Durchführung von Kurz-HTAs. Ihre Methodenbeschreibungen bilden eine weitere wichtige Literaturbasis für dieses Projekt, ebenso wie die methodischen Ausarbeitungen der Arbeitsgruppe 4 des ECAHI Projektes (ECHTA/ECAHI 2001).

Das Schnellverfahren konzentriert sich nur auf die Aspekte der „medizinischen“ Effektivität, eine gesundheitsökonomische Bewertung kann im Rahmen des vorgegebenen Zeitrahmens nicht mit abgedeckt werden.

Eine gleichzeitige Bearbeitung des Themas nach den methodischen Vorgaben für ein vollständiges HTA-Verfahren erfolgte aus zeitlichen Gründen nicht. Dieser sicherlich notwendige Vergleich zwischen einer Kurz- und Langversion wird sicherstellen müssen, dass eine Vorgehensweise, wie in dem vorliegenden Manual beschrieben, zu keiner Verzerrung der Schlussfolgerungen führt.

Allgemeine Aspekte: Entwicklung des Protokolls und „Peer Review“

Protokoll

Bei der Erstellung eines systematischen Review bzw. eines HTA-Berichtes mit integrierter systematischer Literaturübersicht handelt es sich um eine wissenschaftliche Arbeit. Wie für jede wissenschaftliche Untersuchung wird daher auch für die Reviewerstellung gefordert, vor Beginn der Arbeiten ein Studienprotokoll zu erstellen, dessen Vorgaben das Projekt

- inhaltlich bestimmen und abgrenzen,
- die einzusetzende Methodik festschreiben,
- eine Ressourcen- und Zeitplanung vorgeben und
- Maßnahmen bei Problemen und Ablauffehlern vorschlagen.

Die wohl exaktesten Vorschläge zur Protokollentwicklung für systematische Literaturübersichten wurden von der Cochrane Collaboration formuliert:

- Problemstellungen müssen in gut ausformulierten Fragen dargelegt werden. Diese sind zu strukturieren nach Studienpopulation, Intervention, Zielgröße (klinisch und wirtschaftlich) sowie Studiendesign.
- Das Protokoll enthält die Literatursuchstrategie zur Identifikation relevanter Studien, die Ein- und Ausschlusskriterien für die Primärmaterialien, die Qualitätsbeurteilungsparameter für Primärmaterialien (methodische Qualität) und die Methodik der Datenextraktion und der Datensynthese.
- Studienergebnisse können beeinflusst (und somit auch verzerrt) werden durch Faktoren wie Charakteristika der Studienpopulationen, Unterschiede bei der Durchführung von Interventionen (gewollt oder ungewollt) oder unterschiedliche Maße bei der Bewertung von Zielgrößen. Es wird auch diskutiert, ob unterschiedliche Studiendesigns per se eine Beeinflussung der Ergebnisse verursachen können. So wird beispielsweise postuliert, dass positive Ergebnisse von Interventionen in randomisierten Studien weniger deutlich nachgewiesen werden als in nicht-randomisierten Designs (Sterne 2000). In Abhängigkeit von der Fragestellung hat das Reviewprotokoll festzulegen, welche dieser Einflussfaktoren besonders sorgfältig zu analysieren sind.
- Das fertige Protokoll muß von den Verfassern und einer fachlichen und methodischen Expertengruppe genehmigt werden, bevor mit dem eigentlichen Bericht begonnen werden kann.
- Wenn ein Review-Protokoll genehmigt ist, sollte es nur aus guten Gründen modifiziert werden. Solche Abweichungen müssen begründet und schriftlich fixiert werden.

Prinzipiell erfolgt die Protokollerstellung für HTA-Berichte in ähnlicher Weise. Die bisherigen Erfahrungen im deutschen HTA-Projekt haben jedoch gezeigt, dass sowohl der Umfassenheitsanspruch von HTA selber, als auch die (bisher gewollte) weitgehende Freiheit der Wissenschaftler, den Rahmen der Bearbeitung abzustecken, dazu führen, dass sehr breite Fragestellungen festgeschrieben werden, die kaum in einem überschaubaren Zeitrahmen abzuarbeiten sind. Der Schritt der Protokollerstellung ist für ausführliche HTAs teilweise im „Toolkit“ festgeschrieben, wurde aber bei den bisher bearbeiteten HTA-Projekten arbeitsgruppenintern und wenig transparent durchgeführt.

Wir denken, dass sich durch ein validiertes Methodenmanual mit definierten SOPs für einzelne Arbeitsschritte der langwierige Prozess der Abstimmung und Genehmigung des Protokolls strukturieren und vereinfachen läßt.

Peer Review

Im Peer Review-Verfahren soll sicher gestellt werden, dass der HTA-Bericht sowohl methodisch als auch inhaltlich-thematisch dem wissenschaftlichen Erkenntnisstand entspricht („Qualitätssicherung“). Ein weiteres Ziel ist die Erhöhung der Akzeptanz der Berichte durch Einbindung von ausgewiesenen Fachgebietsexperten (s. u.) in das Reviewverfahren. Im Toolkit der German Scientific Working Group findet sich eine Beschreibung des Prozederes für einen zweistufigen Peer Review. In der ersten Stufe des Reviews wird eine vorläufige Endversion des HTA-Berichtes vor allem aus methodischer Perspektive durch 2 - 3 Mitglieder der Arbeitsgruppe selbst begutachtet (*interner Review*). Diese Einzelgutachten werden dem wissenschaftlichen Sekretariat übermittelt, zu einer Stellungnahme zusammengeführt und den Autoren des HTA-Berichtes vier Wochen nach Eingang des Themas zugestellt. Die Verbesserungsvorschläge sollen umgesetzt und eine Version zur Abnahme vorgelegt werden.

Ein sich anschließendes *externes Reviewverfahren* wird im wesentlichen unter inhaltlich-thematischen Gesichtspunkten mit der den Zielen „Qualitätssicherung“ sowie „Erhöhung der Akzeptanz des HTA-Reports gegenüber potentieller Kritik von außen“ durch einen bzw. zwei ausgewiesene Fachexperten, die anhand von Publikationslisten ermittelt wurden, durchgeführt.

Für das Schnellverfahren ist vorgesehen, Fachexperten (zum medizinischen Verfahren, zum betreffenden Krankheits- / Störungsbild, zur Reviewmethodik) und Nutzer (Auftraggeber, andere Entscheidungsträger) bereits in einer sehr frühen Phase der Berichtsbearbeitung einzubinden, nämlich schon bei der Festlegung der präzisen Fragestellung. Diesem eingangs identifizierten Personenkreis wird der abgeschlossene Bericht auch später zum Peer-Review vorgelegt. Ein breiteres Reviewverfahren erübrigt sich auf diese Art und Weise, da durch die vorgegebenen SOPs hohe methodische Qualität und die frühe Einbindung der späteren „Nutzer“ des HTA-Berichtes ein exakter und präziser thematischer Zuschnitt auf die Fragestellung („tailor to demand“) des Auftraggebers begünstigt werden würde.

Die Einführung von standardisierten Arbeitsanweisungen (standardised operating procedures, SOP's) bei der Erstellung von HTA-Berichten ist als ein weiterer Baustein in der Qualitätssicherung vorgesehen. Nach Gibis et al. (2001) befindet sich die Entwicklung von Standards für HTA-Berichte insgesamt in einem sehr frühen und eher explorativen Stadium. Vor einer Nutzung solcher Standards in der täglichen Arbeit der Selbstverwaltung sei sicherzustellen, dass deren Effektivität und Praxistauglichkeit hinreichend geprüft wurde.

Nachfolgend stellen wir die sieben SOP's dar. Ausgangspunkt ist immer die methodische Arbeitsweise für ausführliche Verfahrensbewertungen der deutschen HTA-Arbeitsgruppe. Liegen aus dieser Arbeitsgruppe zu einem Arbeitsschritt keine Empfehlungen vor, bilden publizierte methodische Vorgehensweisen anderer Arbeitsgruppen (z. B. der Cochrane Collaboration) die Basis für unsere Empfehlungen. In diesem Fall wird die entsprechende Publikation direkt zitiert. Anschließend soll das

mögliche Kürzungspotential und seine Umsetzung im laufenden Kurz-HTA-Projekt dargestellt werden. Zum Abschluss erfolgt eine kritische Bewertung und als Schlussfolgerung der Vorschlag einer allgemeingültigen Arbeitsanleitung.

I C.3 Vorschläge für Standardized Operating Procedures (SOPs)

Im Folgenden werden sieben Entwürfe für standardisierte Arbeitsanleitungen vorgestellt. Jede Vorstellung gliedert sich in fünf Teile. Der erste Teil, "Hintergrund", beschreibt die methodische Ausgangslage, d. h. welche Vorgaben existieren für diesen Arbeitsschritt im internen "Toolkit" der deutschen HTA-Arbeitsgruppe (in Grundzügen enthalten in Bitzer et al. 1998) oder in der internationalen Literatur. Unter dem anschließenden Gliederungspunkt "Methoden: Modifikations- / Kürzungspotential" wird die Erstellung der neuen, verkürzten Arbeitsanleitung aus den unter "Hintergrund" beschriebenen Vorgaben beschrieben. Im gleichen Abschnitt wird auch die Arbeitsfassung des SOP präsentiert. Der darauf folgende Abschnitt "Ergebnisse" beschreibt die Erprobung und den Einsatz der neuen Arbeitsvorschrift bei der Bewertung der quantitativen Ultraschallverfahren zur Abschätzung des Risikos für osteoporosebedingte Frakturen. Im Abschnitt "Diskussion" werden Erfahrungen und Konsequenzen aus der konkreten Themenarbeit dargestellt und in den "Schlussfolgerungen", werden letztendlich die, möglicherweise modifizierten, Vorschläge der Arbeitsanweisungen für Schnell-HTA-Verfahren präsentiert.

I C.3.1 SOP 1: Feinabstimmung der Reviewfragestellung mit Nutzern und Fachgebietsexperten

I C.3.1.1 Hintergrund - Status Quo (Ausführliches HTA)

„Rapid assessments“ sollen innerhalb eines überschaubaren Zeitraums valide und verwertbare Informationen für eine konkrete Entscheidungssituation in einem definierten gesundheitspolitischen Kontext liefern. Um eine Fragestellung mit vertretbarem Zeit- und Ressourcenaufwand nach etablierter Reviewmethodik bearbeitbar zu machen, müssen im Vorwege

- die relevanten Fragen im Kontext des Technologieeinsatzes identifiziert und
- nach ihrer Wichtigkeit geordnet (priorisiert) werden.

Die engen Zeitvorgaben können insbesondere dann nicht eingehalten werden, wenn zu viele Aspekte zur Bewertung ausgewählt werden bzw. wenn die gewählten Aspekte nicht den Informationsbedarf der Entscheider treffen und daher Nachbesserungen notwendig machen.

Im Toolkit für Verfasser von HTA-Berichten vom 07.06.1999 wird lediglich eine knappe Darstellung von Anlass und Ziel der Evaluation (Auftraggeber, Zielsetzung, Konsequenzen) unter dem Gliederungspunkt Policy Question gefordert. Die im deutschen HTA-Projekt bisher bearbeiteten Themen wurden zum überwiegenden Teil unter methodisch wissenschaftlichen Gesichtspunkten gewählt und nicht, weil sie einen realen Informationsbedarf decken sollten. Der Druck, Informationen zu einem

bestimmten Entscheidungstermin bereit zu stellen, war somit nicht vorhanden. Auch die Auswahl der zu bearbeitenden Subfragestellungen erfolgte in erster Linie aus wissenschaftlichem oder persönlichem Interesse. Dementsprechend knapp sind die Vorgaben zur Themenfindung / -abstimmung im derzeit noch gültigen „Toolkit“ gehalten, sie werden nachfolgend aufgelistet und kommentiert:

- Eine speziell ausgearbeitete Prozedur zur Identifikation der dringlichsten Fragestellungen für einen HTA-Bericht ist zum jetzigen Zeitpunkt noch nicht festgeschrieben.
- Anlass und Ziel der Bewertung (Auftraggeber, Zielsetzung, Konsequenzen) sind in einer ca. einseitigen „Policy Question“ darzulegen.
- Die konkreten „Forschungsfragen“ (ca. 5 pro Review) werden aus der Darstellung des „Hintergrundes“ abgeleitet. Hierzu sind vorbereitend darzustellen: die Technologie mit Funktionsprinzip und technischen Beschreibungen, Standards, Zulassungsmodalitäten, Qualitätskontrollen, Wartungsbedarf sowie praxisrelevante Aspekte wie Einbindung in Arbeitsabläufe, Zeit, Zusatzausrüstung und Ausbildungsbedarf. Eine Beschreibung der „Krankheitslast“ erfolgt aus Informationen zur Epidemiologie, Pathophysiologie, zu Risikofaktoren, Klinik, Verlauf sowie diagnostischen und therapeutischen Ansätzen. Sozio-ökonomische Gesichtspunkte (z. B. für Patienten, Kostenträger, Gesellschaft) sollen ebenso erwähnt werden wie Ziele, Ansprüche und mögliche Alternativen des Technologieeinsatzes im deutschen Systemkontext. Hauptinformationsquellen für diesen Gutachtenteil bilden Bücher und Übersichtsarbeiten, (administrative) Statistiken aber auch Expertenkontakte. Es wird besonders Wert auf die Formulierung von präzisen, im Bericht zu beantwortenden Fragestellungen gelegt, jedoch wird das Prozedere hierzu im Toolkit nicht weiter ausgeführt.
- Die Einbindung von Experten und Nutzern in die Themenfindung ist im Standardverfahren nicht vorgesehen – mit der Konsequenz, dass viele Berichte einen unverhältnismäßig hohen Ressourcenverbrauch hatten, bei vielen Berichten im Laufe der Bearbeitung weitere Einengungen der Fragestellungen erforderlich wurden, einige Berichte nach dem externen Expertenreview komplett überarbeitet werden mussten und schließlich einige Gutachten ohne Beachtung durch Fachkreise „im Schrank“ verschwanden.

I C.3.1.2 Methoden: Modifikations- / Kürzungspotential

Für das SOP 1 ist eher von Modifikations- als von Kürzungspotential zu sprechen, da es unter den Verfahrensvorschlägen im „Toolkit“ zu diesem Punkt keine abgestimmte Vorgehensweise gibt. Die hier vorgeschlagene kann also nur aufwendiger werden, wird sich unseres Erachtens aber in einer erheblichen Zeitersparnis, größeren Zielgenauigkeit und Bedarfsgerechtigkeit sowie höherer Akzeptanz der Bewertungen auswirken.

Die SOP umfasst:

- Akquisition und Sichtung von Hintergrundinformationen zu Technologie und „Zielkondition“;
- Auswahl eines „Expertenkreises“ (Nutzer, Methodiker, Technologieexperten) zur Unterstützung der Themenfindung (und für den Peer Review);
- Formulierung weniger relevanter Fragestellungen;
- Abstimmung der Fragestellungen im modifizierten Delphi-Verfahren.

Die Präzisierung der relevanten Fragestellungen kann zu einer Reduktion der zu bearbeitenden Informationsfülle führen und somit zu einer möglichen Verkürzung der Berichtsbearbeitung. Hierfür ist es erforderlich, die Fragestellung möglichst stark einzugrenzen, sie indikationsspezifisch zu formulieren und in Absprache mit den Fachexperten und Nutzern/Abnehmern des Kurz-HTAs zu konkretisieren. Die Fokussierung auf wenige gezielte Fragestellungen beschleunigt den Bearbeitungsprozess.

Der im entscheidungsrelevanten Kontext präzierte Fragenkatalog sollte durch diesen weit gefassten Expertenkreis in einem schriftlichen Delphi-ähnlichen Verfahren abgestimmt werden, damit sichergestellt wird, dass keine relevanten Informationen verlorengehen und die Zielsetzungen des HTA-Berichtes mit denen der zukünftigen Nutzer im Gesundheitswesen übereinstimmen. Dieses erweiterte Vorgehen kürzt das Gesamtverfahren vermutlich ab, da die sorgfältig ausgesuchten Kooperationspartner schriftlich und damit zeitsparend Stellung nehmen können.

I C.3.1.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt

In unserem Methodenprojekt wurden parallel zwei Arbeitsschritte durchgeführt:

- Zusammenstellung von Hintergrundinformation anhand von Expertenkontakten und einer orientierenden Literaturrecherche,
- Identifikation der präzise zu bearbeitenden Fragestellungen durch Einbeziehung von Kooperationspartnern mit Hilfe des Delphi-Verfahrens.

Die Aquisition von Hintergrundinformationen zu Technologie und Krankheits- / Störungsbild

Um Zugang zum Hintergrundwissen zu erhalten, muss im Vorfeld der HTA-Berichterstellung eine orientierende Literaturrecherche (Identifikation von Primärliteratur) erfolgen. Diese führten wir in der Medline-Datenbank sowie in der Cochrane Library mit den Schlagwörtern „Osteoporosis“ und „Ultrasound“ durch. Im Internet wurden unspezifische „Suchmaschinen“ (z. B.: www.google.de) und die gleiche Suchstrategie eingesetzt.

Unser Schwerpunkt lag zunächst auf der Recherche nach Übersichtsartikeln und Reviews aus den Jahren 1990 bis 2001, um abzuklären, ob die Fragestellung nicht schon ausreichend verlässlich in bereits vorhandenen systematischen Reviews behandelt worden war. Darüberhinaus sollten aktuelle Reviews (Jahrgänge 1997 – 2001) die notwendigen Hintergrundinformationen zu Technologie und Zielkondition liefern.

Insgesamt fanden sich 35 Übersichtsarbeiten aus den verschiedensten Ländern und von unterschiedlichsten Autorengruppen sowie ein kanadischer HTA-Bericht aus dem Jahre 1998 über quantitative Ultraschallverfahren zur Knochendichtemessung.

Die allesamt unsystematischen Reviews kamen zum überwiegenden Teil aus den USA, gefolgt von Großbritannien, Deutschland und Frankreich sowie vereinzelt aus Kanada, Australien, den Niederlanden und Belgien.

Diese Übersichtartikel lieferten sehr gute Hintergrundinformationen zur Technologie der quantitativen Ultraschallverfahren, zum derzeitigen Entwicklungsstand (life-cycle) und teilweise auch zu Preisen (Preisniveau der nicht-erstattungsfähigen Leistung, Gerätepreise etc.), zur Pathophysiologie der Osteoporose, zu epidemiologischen Fragestellungen, zu Risikofaktoren und Risikogruppen, zu möglichen Alternativen und insbesondere zum Goldstandard, d. h. dem DXA-Verfahren. Wir fanden keine Angaben zu Frequenz und Modalitäten des Technologieeinsatzes im deutschen Systemkontext.

Darüberhinaus wurden verschiedene medizinische Fachgesellschaften online kontaktiert (Fachgesellschaften der Orthopädie, Osteologie, Gynäkologie, Innere Medizin etc.); auf diese Art und Weise konnten wir aktuelle Informationen zu Indikationsregeln im jeweiligen Fachbereich sammeln.

Zusammenfassend kann hervorgehoben werden, dass die erforderlichen Hintergrundinformationen ohne Probleme aufgefunden werden konnten. Lediglich Zahlen bzw. Angaben über die Verbreitung der QUS-Verfahren und ihre Akzeptanz als selbst zu finanzierende (IGEL-)Leistung durch die Patienten waren zunächst nicht zugänglich. Hierzu wurden auf einem Fachkongress der Orthopäden und Unfallchirurgen diejenigen Firmen und Unternehmen am Messestand aufgesucht, die Quantitative Ultraschallgeräte herstellen. In einem persönlichen Interview wurden dann Verkaufszahlen, Kosten und Abrechnungsmöglichkeiten erfragt.

Die Auswahl eines „Expertenkreises“ (Nutzer, Methodiker, Technologieexperten) zur Unterstützung der Themenfindung (und für den Peer Review)

Anhand der Literatur konnten für den deutschsprachigen Raum mehrere wissenschaftliche Arbeitsgruppen identifiziert werden, die sich vorrangig mit der Ultraschall-diagnostik des Knochens beschäftigen, hierunter vor allem die Abteilung für diagnostische Radiologie der Universität Kiel.

Zusätzlich wurden weitere Fachexperten zum Thema Osteoporose eingebunden: die deutsche EVOS (European Vertebral Osteoporosis Study) Arbeitsgruppe und die DAGO (Deutsche Arbeitsgemeinschaft Osteoporose).

Nutzer eines Kurz-HTA sind Auftraggeber oder andere Entscheidungsträger auf gesundheitspolitischer Ebene, die u. a. auch gutachterlich zur Wirksamkeit und Wirtschaftlichkeit medizinischer Technologien Stellung nehmen müssen. Grundsätzlich in Frage kommen: Ärzteschaft (standes- und berufsrechtliche Organisationen); Spitzenverbände der Kranken- und Rentenversicherungsträger (auf Bundes- und Landesebene), Krankenkassen und medizinischer Dienst (auf Bundes- und Landesebene); Klinische Forschungsinstitutionen (Universitätskliniken mit den wichtigsten klinischen Bereichen); Theoretische Forschungsinstitutionen (universitäre und außeruniversitäre wissenschaftliche Institute, HTA-Einrichtungen); Wissenschaftliche Fachgesellschaften; Ausgewählte Patienten- und Verbraucherorganisationen; Gesundheitspolitik auf Bundes- und Landesebene. In unserem Methodenprojekt wurden Personen aus folgenden Institutionen um Mitarbeit gebeten: KBV in Köln, MDK in Schleswig-Holstein und MDS in Essen.

Die methodisch arbeitenden Mitarbeiter aus der Universitätsabteilung für Epidemiologie, Sozialmedizin und Gesundheitssystemforschung in Hannover und dem DIMDI bilden zusammen mit den Fachexperten und den potentiellen Nutzern den „Expertenkreis“.

Formulierung weniger, relevanter Fragestellungen

Nach Durchsicht der orientierenden Literaturrecherche und der verfügbaren Hintergrundinformationen wurde eine kurze Darstellung des entscheidungsrelevanten Kontexts ("Policy Question" vergl. Kapitel II.C.1) verfasst und fünf Fragen formuliert, die uns essentiell zur Beschreibung des Stellenwertes von Ultraschallverfahren im Kontext der Osteoporoseversorgung schienen. Ihre Beantwortung sollte daher den Kern des Kurz-HTA-Berichtes bilden.

1. Wie korrekt läßt sich das Frakturrisiko mit Hilfe der quantitativen Ultraschallverfahren abschätzen (relatives Frakturrisiko im Vergleich zu Personen mit Normalwerten)?
2. Welches der unterschiedlichen QUS-Verfahren ist hierzu am besten geeignet?
3. Welche qualitätssichernden Maßnahmen sind vor einem Einsatz in der ärztlichen Praxis zu realisieren, um zu validen Ergebnissen zu kommen?
4. Gibt es weiteren Forschungsbedarf zur Klärung des Stellenwertes der Ultraschalldiagnostik im Rahmen der Osteoporoseversorgung?
5. Wie wären die ermittelten Ergebnisse zur Patienteninformation und Entscheidungshilfe nutzbar?

Abstimmung der Fragestellungen im modifizierten Delphi-Verfahren

Die Delphi-Technik (Delphi-Verfahren, Delphi-Konferenz), die bereits 1948 entwickelt wurde, ist eine interaktive Umfragemethode. Charakteristisch sind das kontrollierte Feedback und die Befragung wechselseitig anonymisierter Teilnehmer. Ablauf: (1) Einholung anonymisierter Meinungen von Experten mit Hilfe eines Fragebogens oder eines Interviews. (2) Durchführung mehrerer Befragungsrunden, wobei nach jeder Runde die eingetroffenen Antworten zusammengefasst und den Befragten erneut zur Begutachtung zugeschickt werden. So kommt es zur systematischen Modifikation und Kritik der zusammengefassten anonymen Antworten. (3) Erreichung einer Gruppenantwort durch eine Zusammenfassung (häufig statistisch) der individuellen Meinungen in einer Abschlussrunde. Das Delphi-Verfahren sollte abgeschlossen werden, wenn eine Konvergenz der Meinungen erkennbar wird oder wenn das Interesse der Teilnehmer spürbar abnimmt.

Kennzeichnend für das Verfahren sind also mehrere Befragungsrunden, in denen die Experten immer wieder mit ihren Aussagen und denen ihrer Kollegen konfrontiert werden.

Zusammen mit der "Policy Question", einer knappen Technologiebeschreibung und der Darstellung der Zielgröße (Risiko osteoporotischer Frakturen) haben wir den oben beschriebenen Fragenkatalog im Rahmen eines modifizierten, schriftlichen Delphi-Verfahrens an den bekannten Expertenkreis geschickt. Es wurde nur eine Befragungsrunde durchgeführt, sowie nach Zusammenfassung der unterschiedlichen Meinungen und Kommentare eine Abschlussrunde.

Die Kooperationspartner wurden um Stellungnahme und konstruktive Kritik sowohl zum Hintergrund, als auch zu den präzisen Fragestellungen gebeten, wobei jeder Experte Gelegenheit hatte, Kommentare aus seiner spezifischen Perspektive einzubringen.

Technologie:

- Ist das Verfahren korrekt beschrieben?
- Entsprechen die formulierten Fragestellungen dem wissenschaftlichen Erkenntnisstand zum Verfahren?

Krankheit:

- Ist die Darstellung der zu beeinflussenden Zielkondition korrekt?
- Ist der (derzeitige) Einsatz des Verfahrens korrekt dargestellt?
- Sind die formulierten Fragestellungen relevant für den Informationsbedarf des Anwenders?

Potentielle Nutzer:

- Ist die Entscheidungssituation richtig erkannt?
- Sind die formulierten Fragestellungen relevant für den Entscheider?

Methodiker:

- Werden die formulierten Fragestellungen mithilfe einer Literaturanalyse für beantwortbar gehalten?

In dem von uns durchgeführten Delphi-Verfahren wurden insgesamt acht Personen bzw. Institutionen angeschrieben.

Verfahrensexperten	n = 1
Osteoporoseexperten	n = 2
Potentielle Nutzer	n = 3
Methodenexperten	n = 2

Insgesamt haben n = 4 (3 potentielle Nutzer, 1 Methodiker) zeitgerecht, d. h. innerhalb einer 4-Wochen-Frist geantwortet. Ihre Kommentare sind stichwortartig zusammengefasst:

Tabelle 3: Modifikationsvorschläge aus der Expertenbefragung

Kommentare
Frage 1 zu unpräzise und missverständlich formuliert (3 Kommentare)
Klare Abgrenzung zur Validitäts-/Reliabilitätsfeststellung (Aufnahme in die Hintergrunddarstellung) erforderlich
Klare Abgrenzung zur „Osteoporosediagnostik“ (d. h. hier geht es um die Erfassung des Frakturrisikos, nicht um die Beschreibung der Knochen“dichte“ und damit Diagnosestellung nach WHO) erforderlich
Vorschläge für zusätzlich zu bearbeitende Fragestellungen
Systematische Statusbestimmung, inklusive Kostenanalyse (2 Kommentare)
Therapeutische Konsequenzen eines (niedrigen) QUS-Befundes
Eignung des Verfahrens zur Therapieüberwachung
Vergleichende Bewertung zu anderen Verfahren der „Knochendichtemessung“
Stratifikation der Bewertung nach Osteoporosetyp (I,II,sekundäre Formen)
Stratifikation der Bewertung für die Bereiche Prävention und Kuration
Vorschläge zur Reduktion des Bearbeitungsumfanges
Aspekte Qualitätssicherung, Patienteninformation und Forschungsbedarf zurückstellen

Diese Kritikpunkte, Verbesserungsvorschläge und Einwände wurden von uns berücksichtigt und eingearbeitet und der Katalog erneut zirkuliert. Nach Zustimmung

durch die Kooperationspartner wurde die überarbeitete Fassung als Grundlage zur weiteren Berichtsbearbeitung übernommen.

Im folgenden findet sich der von uns korrigierte Fragenkatalog, der schlussendlich im HTA-Schnellverfahren bearbeitet wurde:

1. Wie valide läßt sich das Frakturrisiko mit Hilfe der quantitativen Ultraschallverfahren im Vergleich zum DXA-Verfahren abschätzen (prädiktive Validität: relatives Frakturrisiko im Vergleich zu Personen mit Normalwerten)?
2. Welches der unterschiedlichen QUS-Verfahren ist hierzu am besten geeignet?
3. Sollte die Analyse ergeben, dass die quantitative Ultraschalldiagnostik eine valide und zuverlässige Abschätzung des Frakturrisikos erlaubt: welche qualitätssichernden Maßnahmen sind dann vor einem Einsatz in der ärztlichen Praxis zu realisieren, um zu reliablen und unverzerrten Ergebnissen zu kommen?
4. Welchen weiteren Forschungsbedarf zur Klärung des Stellenwertes der Ultraschalldiagnostik im Rahmen der Frakturprävention gibt es?
5. Wie wären die unter 1 – 3 ermittelten Ergebnisse zur Patienteninformation und zum Verbraucherschutz zu nutzen?

I C.3.1.4 Diskussion: Bewertung, eigene Erfahrungen

Bei der orientierenden Literaturrecherche und Sichtung kann man durch erfahrene Mitarbeiter viel Zeit einsparen. Der Zugang zu den relevanten Datenbanken ist größtenteils kostenlos, über DIMDI lassen sich MEDLINE und Healthstar anwenderfreundlich recherchieren. Die orientierende Suche zu Hintergrundinformationen bei den Fachgesellschaften und über Suchmaschinen gestaltete sich problemlos.

Um die Orientierungsphase möglichst kurz zu halten, wäre es unserer Meinung nach wünschenswert, einen entsprechenden Kreis von Experten der verschiedensten Fachrichtungen regional und überregional zu etablieren, die je nach Erfordernis in unterschiedliche HTA-Projekte eingebunden werden könnten. Durch eine größere Anzahl von Kooperationspartnern wird sichergestellt, dass keine relevanten Informationen verlorengehen und die Ziele und Fragestellungen des HTA-Berichtes für die Nutzergruppe entscheidend und bedarfsorientiert sind. Hierzu ist eine enge Zusammenarbeit von Auftraggebern und –nehmern bei der Fokussierung der Fragestellung und der darauf exakt ausgerichteten Konzeption des HTA-Berichtes erforderlich.

Kritisch abzuwägen ist allerdings, dass erfahrungsgemäß nicht alle Kooperationspartner zeitgerecht antworten werden.

Der Rundbrief mit den für uns relevanten Fragen wurde an alle von uns vorher ausgewählten „Partner“ geschickt; innerhalb einer 4-Wochen-Frist haben lediglich die potentiellen Nutzer geantwortet, nach sechs Wochen dann auch einer der beiden Methodiker. Da der Fragenkatalog bzw. die Policy Question die Basis zur Erstellung

eines HTA-Berichtes darstellt, wird sich das Verfahren automatisch verzögern, wenn nicht im vorgegebenen Zeitrahmen geantwortet wird.

I C.3.1.5 Schlussfolgerung: SOP 1

Vorschlag für eine generalisierbare Arbeitsanleitung:

1. Vorbereitung: Hintergrundinformationen einholen

- Zur Verfügung gestellte Schlüsselpublikationen
- Orientierende Literaturrecherche: aktuelle Übersichtsarbeiten
- Expertenkontakte (Technologie/Zielkrankheit)
- Herstellerinformationen
- Informationen aus Routinedatenbeständen (z. B. Verschreibungsverhalten)

2. Bildung des Expertenkreises aus den Gebieten*:

- Technologie
- Krankheit
- Potentielle Nutzer
- Methodik

* In Einzelfällen auch weitere Spezialkompetenzen

3. Leitfragen zur Formulierung der "Policy Question" (Entscheidungskontext):

- Warum ist eine Bewertung der Technologie sinnvoll?
- Warum ist die Bewertung zum jetzigen Zeitpunkt sinnvoll?
- In welche Entscheidungen soll das Ergebnis des Berichtes einfließen?
- Welche grundlegenden Probleme kann die Verfahrensbewertung klären?

4. Ableitung weniger, präziser, zentral wichtiger Fragen

5. Modifiziertes email-basiertes Delphi-Verfahren zur Fragenabstimmung

- Zirkulieren der Policy Question und nachgeordneten Fragestellungen mit Antwortfristen im Expertenkreis
- Entgegennahme von Rückmeldungen
- Einarbeiten der Kommentare
- In zweiter Runde Verabschiedung des Fragenkataloges

I.C.3.2 SOP 2: Literaturrecherche für HTA-Schnellverfahren

I.C.3.2.1 Hintergrund: Status Quo (Ausführliches HTA)

Zentraler Bestandteil aller ausführlichen Verfahrensbewertungen (HTA-Berichte) ist eine systematische Literaturübersicht. Abhängig von der präzisen Fragestellung liefert die Übersicht den wissenschaftlichen Kenntnisstand, beispielsweise zur Wirksamkeit oder zur Sicherheit des Verfahrens bzw. zu erforderlichen Qualitätssicherungsmaßnahmen als Voraussetzung für den Technologieeinsatz. Per definitionem ist die systematische Literaturübersicht bestrebt, einen wissenschaftlichen Kenntnisstand umfassend und verzerrungsfrei abzubilden. Eine umfassende Literaturrecherche muss somit die Basis der Übersichten bilden.

Ziel dieser Recherchen im Rahmen von ausführlichen HTA-Bewertungen ist die Ermittlung aller abgeschlossenen oder laufenden Studien zu den Fragestellungen des HTA-Vorhabens. Als typische systematische Fehler (=bias) in Literaturübersichten führen der *Publikationsbias*, der *Language-Bias* und der *Retrieval-Bias* zu Ergebnisverzerrungen. Durch sorgfältige Recherchestrategien können jedoch diese Fehler erkannt und die Ergebnisse der Reviews entsprechend diskutiert werden. Für die Arbeit der „German Scientific Working Group....“ wurden zwei unterschiedlich aufwendige Vorgehensweisen bei der Literaturrecherche dokumentiert.

1. Im „Toolkit“ schlägt die Projektgruppe Hannover - aufgrund ihrer Erfahrungen bei der Suche und Beschaffung der Literatur - ein praxisorientiertes Verfahren für eine systematische und schnelle Vorgehensweise vor. Es werden Empfehlungen zu elektronischen Datenbankrecherchen aber auch zu weiterführenden Strategien („Hand-searching“, Expertenkontakte) abgegeben.

Die Hinweise für die elektronischen Recherchen beinhalten:

- die zu berücksichtigenden Datenquellen,
- die Suchstrategien in den Datenbanken,
- die Dokumentation der Literaturrecherche,
- die Literaturbeschaffung.

Im Anhang werden die wichtigsten wissenschaftlichen Datenbanken beschrieben und der jeweilige Internet-Zugang aufgelistet.

Die Basissuche für Primärstudien sollte in MEDLINE und EMBASE erfolgen; für die Suche nach systematischen Reviews sollte die Cochrane Library an erster Stelle stehen.

Darüberhinaus werden noch eine Reihe von kleineren speziellen Datenbanken (z. B. HealthStar, Cancerlit, SOMED etc.) aufgeführt, die, in Abhängigkeit von der aktuellen Thematik, für die Literaturrecherche genutzt werden sollten.

Zusätzlich zu den Literaturdatenbanken sollten Kongressbände der wichtigsten Kongresse für das jeweilige Thema nach relevanten Studien durchsucht werden. Hierzu gehören Weltkongresse, europäische, nordamerikanische und deutsche Kongresse. Auch wird gefordert, wichtige Journals von Hand zu durchsuchen, um in den letzten Wochen erschienene Publikationen berücksichtigen zu können.

Wichtig ist weiterhin der frühzeitige Kontakt zu Experten des jeweiligen Fachgebietes bzw. zu den Fachgesellschaften, um laufende Studien zum jeweiligen Thema in Erfahrung zu bringen. Abschließend sollte eine ausführliche Dokumentation der Literaturrecherche erfolgen, um eine Überprüfung und Reproduzierbarkeit (für spätere Suchen bei z. B. Updates) zu gewährleisten.

Eine ausführliche aktuelle Darstellung dieser Vorgehensweise wurde in der Zeitschrift für Ärztliche Fortbildung und Qualitätssicherung im August 2001 (Droste et al. 2001) veröffentlicht. Ergänzend zu den obigen Ausführungen finden sich in der Publikation Angaben zum Recherchezeitraum, der im allgemeinen für ausführliche HTAs mit rund 10 Jahren (retrospektiv) angegeben wird. Bei einigen Themen – insbesondere bei neueren Technologien – kann dieser Zeitraum jedoch deutlich kürzer, bei anderen Themen (ältere Technologien) auch länger sein.

2. Als Weiterentwicklung der oben beschriebenen Arbeitsanleitungen wurden unter Federführung der HTA-Arbeitsgruppe am DIMDI für das Vorgehen bei der Literaturrecherche zwei „Standard Operating Procedures“ (SOPs) vorgeschlagen. Die SOPs gelten für die Erstellung von HTA-Berichten zur medizinischen Effektivität einer Technologie innerhalb der deutschen HTA-Arbeitsgruppe (SOP Datenbasis und SOP Literaturrecherche). Die beiden SOPs sind (als bisher einzige) in der Datenbasis des Informationssystems „Gesundheitsökonomische Evaluation/Health Technology Assessment“ dokumentiert. Die Standards für die Erstellung von HTA-Berichten werden vom DIMDI (Deutsches Institut für Medizinische Dokumentation und Information) aktuell gehalten und sind im Internet unter www.dimdi.de/germ/evalua/sopstart.htm abzurufen.

I C.3.2.2 Methoden: Modifikations- / Kürzungspotential

Der im HTA-Bericht zusammengefasste Stand der wissenschaftlichen Kenntnisse muss den Anforderungen nach Vollständigkeit und Transparenz genügen. Es ist daher notwendig eine umfassende Literaturrecherche durchzuführen, die in der Regel sehr viel Zeit (mehrere Monate, inklusive Lieferzeiten für Fernleihbestellungen) in Anspruch nimmt. Die Literaturrecherche gehört damit zu den Arbeitsschritten, die im Gesamtkonzept den größten Zeitaufwand beanspruchen. Die Methodik eines Schnellverfahrens sollte daher im Abschnitt Literaturrecherche erheblich Zeit einsparen, ohne systematische Verzerrungen in die Auswahl der Basisliteratur einzuführen.

Die hier angeführten Arbeitsschritte sollen die bei HTA-Verfahren gebräuchliche systematische Herangehensweise an die Recherche bewahren und eine absolut transparente Dokumentation sicherstellen.

Mögliche Verkürzungsvorschläge für die **Literaturrecherche**:

- Hierarchisierung nach Publikationstypen und Eingrenzung der Datenbankauswahl
- Eingrenzung des Publikationszeitraumes (z. B. auf < 10 Jahre)

Unserer Meinung nach liegt das Hauptkürzungspotential in der hierarchischen Durchführung der Literaturrecherche und der Begrenzung der Recherchen in den biomedizinischen Datenbanken, da hierdurch die Zahl der recherchierten und damit zu bewertenden Dokumente zumeist deutlich reduziert werden kann. Vor allem systematische Übersichten nach dem Cochrane-Standard können zeitaufwendige Auswertungen von Primärstudien entbehrlich machen. Diese Publikationen, die systematische Informationszusammenfassungen enthalten (z. B. HTA-Berichte, Reviews, Meta-Analysen) sollen in der Cochrane Library, in den HTA-Datenbanken (NHS CRD Datenbanken: DARE, NEED und HTA, HSTAT, ISTAHC) sowie diversen Leitliniendatenbanken und den Publikationslisten der weltweit tätigen HTA-Institutionen gesucht werden. Eine Liste mit den Charakteristika der zu verwendenden Datenbanken mit Inhalten, Rechercheprinzip, Zugängen, Kosten usw. finden sich im Anhang.

Primärstudien sollen nur zur Abdeckung der „Lücke“ zwischen Redaktionsschluss für die Übersichten und der Gegenwart herangezogen werden oder wenn keine systematischen Übersichten aus allerjüngster Zeit aufgefunden werden können.

Die Wahl der Datenbanken ist themenabhängig und kann deshalb durch die hier vorgeschlagenen SOPs nicht festgelegt werden.

Bei sehr neuen Technologien kann es im Einzelfall hilfreich sein, das britische NHSC (National Horizon Scanning Center) oder Euroscan (The European Information Network) zu kontaktieren. Aus prospektiven Studienregistern, die allerdings zur Zeit noch keinen befriedigenden Ausbaustand aufweisen, lassen sich Hinweise auf laufende Untersuchungen gewinnen.

Eine weitere Beschleunigung des Bearbeitungsprozesses erfolgt, wie üblich, durch die standardisierte Anwendung von Ausschlusskriterien.

Einschränkung der Publikationssprache

Eine weitere Möglichkeit zur Verkürzung der Recherche besteht darin, nur englisch- und deutschsprachige Publikationen zu berücksichtigen. Nach einer Studie von Egger et al. (1997) kann der Ausschluss nicht-englischsprachiger Publikationen in Meta-Analysen zu Verzerrungen („language bias“) führen. Sie fanden heraus, dass Autoren randomisierte Studien mit einem statistisch signifikanten Ergebnis bevorzugt in englischsprachigen Zeitschriften publizieren, während Arbeiten ohne signifikantes

Ergebnis eher in Zeitschriften anderer Landessprachen erscheinen. Egger et al. vermuten eine systematische Verzerrung von Reviewergebnissen hin zu deutlicheren Erfolgen bei alleiniger Berücksichtigung englischsprachiger Publikationen. Eine Untersuchung von Gregoire et al. (1995) dagegen ergibt keinen Hinweis auf relevante Ergebnisveränderungen durch linguistische Einschränkungen. Eine neuere Untersuchung von Moher et al. (2000) ergab ebenfalls keinen Hinweis darauf, dass der Ausschluss nicht-englischsprachiger Studienpublikationen zu verzerrten Ergebnissen führt. Vor diesem Hintergrund scheint es uns vertretbar, für das Schnellverfahren die Einschränkung auf die Publikationssprachen Englisch und Deutsch vorzuschlagen. (Eine Ausnahme würden natürlich Themen bilden, deren Forschungsschwerpunkte im nicht-englisch- oder deutschsprachigen Raum liegen).

Verzicht auf manuelle Recherchen

Eine umfangreiche Untersuchung von Langham et al. (1999) hat dargelegt, dass bei Verzicht auf „hand searching“ und Beschränkung auf eine Medline-Recherche etwa ein Drittel der randomisierten Studien aus den 14 dort betrachteten Zeitschriften übersehen wurden. Allerdings konnte hierdurch nicht die Frage geklärt werden, ob dadurch die Ergebnisse systematischer Reviews nennenswert beeinflusst wurden. Sollte sich im Rahmen der Projektvorarbeiten herausstellen, dass es sich um ein sehr schlecht verschlagwortetes Thema handelt (z. B. bei sehr neuen Technologien), kann es in Ausnahmefällen allerdings auch im Rahmen von Kurz-HTAs sinnvoll sein, zentral relevante Publikationsorgane für ausgewählte Jahrgänge von Hand zu durchsuchen.

Verzicht auf die Suche nach „grauer Literatur“

Unter „grauer Literatur“ werden Studien verstanden, deren Veröffentlichungen nur begrenzte Verbreitung haben, nicht in Recherchesystemen enthalten oder dort falsch verschlüsselt sind, sowie unpublizierte Studien (z. B. Tagungsabstracts, unpublizierte Studienergebnisse, interne Berichte von Pharmafirmen oder Doktorarbeiten). Mc Auley et al. (2000) untersuchten den Einfluß dieser grauen Literatur auf die Ergebnisse von Meta-Analysen und kamen zu dem Schluss, dass das Ergebnis bei Ausschluss grauer Literatur signifikant optimistischer war als bei Einschluss grauer Literatur. Nach Ziegler et al. (2001) haben sich jedoch bei genauerer Betrachtung die Ergebnisse der Meta-Analyse quantitativ nur unwesentlich verändert.

Verzicht auf das Aufsuchen unpublizierter und laufender Studien

Bei Arzneimittelstudien wird z. B. empfohlen auch Experten und Hersteller zu kontaktieren und nach weiteren ggf. unpublizierten Studien oder noch nicht abgeschlossenen Studien zu forschen (Bucher 1998; Cook et al. 1995; Koch et al. 2000).

Der Verzicht auf „hand searching“, grauer Literatur und Experten- und Herstellerkontakte zum Auffinden von unpublizierten oder noch nicht abgeschlossenen Studien beschleunigt die Recherche erheblich. Inwieweit sich hierdurch eine Verzerrung der Ergebnisse ergibt müsste durch vergleichende Untersuchungen abschließend geklärt werden. Die oben erwähnten Untersuchungen von Langham et al. (1999) und McAuley et al. (2000) konnten diese Frage nicht beantworten.

Mögliche Verkürzungsvorschläge für die Such- und Selektionsstrategie:

- Enge Definition von inhaltlichen Ein- und Ausschlusskriterien
- Standardisierte Anwendung von Ausschlusskriterien

Die Einschlusskriterien für Studien sollten so eng wie möglich definiert werden, sowohl unter inhaltlichen als auch unter methodischen Gesichtspunkten. Das Einbringen von Qualitätsanforderungen an die Studien als Einschlusskriterium ist geeignet, den Umfang des Rechercheergebnisses erheblich zu reduzieren (s. auch SOP 3). Hierzu gehören z. B. der routinemäßige Ausschluss von „Case-reports“, „Letters“, „Editorials“, „animal-studies“ und eindeutig unsystematischen Reviews.

- Vor- bzw. Erstselektion anhand von Abstracts

Dieser Punkt bezieht sich auf den Umgang mit den Ergebnissen der elektronischen Literaturrecherchen. Aus den Datenbanken werden in der Regel als Rechercheergebnis bibliographische Angaben der Publikationen und, falls vorhanden, die Kurzfassung erhalten. Es gibt jetzt prinzipiell drei Möglichkeiten, wie mit diesem Ergebnis weiter verfahren werden kann:

1. Anhand dieser Kurzfassungen lässt sich überprüfen, ob eine Publikation den Ein- und Ausschlusskriterien entspricht, sodass ihre Ergebnisse in der Bewertung berücksichtigt werden sollten. Nur diese Arbeiten werden in ihrer Vollversion beschafft und weiter bearbeitet. Artikel, aus deren Kurzfassung keine eindeutige Zuordnung zu Ein- oder Ausschluss hervorgeht bzw. die keine Kurzfassung haben, werden aussortiert.
2. Die zweite Möglichkeit geht zunächst vor wie unter Punkt 1 beschrieben, die Arbeiten mit unklaren oder fehlenden Angaben werden jedoch im Volltext besorgt, gelesen und danach über Ein- oder Ausschluss entschieden.
3. Die dritte Möglichkeit sieht grundsätzlich eine Auswahl anhand der Volltexte vor.

Für das Vorgehen im Rahmen des Kurz-HTAs schlagen wir ein Prozedere wie unter Punkt 1 skizziert vor. Eine Anschlussstudie muss allerdings versuchen sicher zu stellen, dass auf diese Weise keine systematische Verzerrung in die Reviewergebnisse hineingetragen wird, d. h., das Aussortieren von Publikationen mit uneindeutigen Kurzfassungen darf sich nicht auf das Reviewergebnis auswirken.

Zur Reviewmethodik „lege artis“ gehört es, die Selektion und die Entscheidung über den Einschluss von Studien durch zwei unabhängige Personen vorzunehmen und anschließend Diskrepanzen gemeinsam zu diskutieren. Der Verzicht auf die zweite Person spart personelle Ressourcen und durch die entfallende Klärung von Diskrepanzen auch Zeit ein. Selbstverständlich schliesst diese Vorgehensweise die Einholung einer Zweitmeinung in zweifelhaften Situationen nicht aus.

Infrastrukturelle Voraussetzungen

Die Nutzung adäquater und komfortabler Datenbanksysteme zur Recherche und zur elektronischen Dokumentbeschaffung erfordert eine großzügige Mittelausstattung der Kurz-HTA-Projekte. Vor allem die Literaturlieferungen stellen einen nicht unerheblichen Kostenfaktor im Rahmen des Gesamtvorhabens dar.

Bestandskonferenz

Zum Abschluss der Hauptrecherchephase wird ein Arbeitsgruppentreffen (intern, bei vorhersehbar unklarer Sachlage unter Hinzuziehung des Methodikexperten des Projektbeirates) einberufen werden. Diese Besprechung soll, in Kenntnis der Evidenzlage zur Fragestellung, das weitere methodische Vorgehen im Projekt festlegen. Im Prinzip geht es um die Feststellung, welche der drei folgenden Konstellationen in der laufenden Projektarbeit angetroffen wird:

A) Ausreichend Literatur: mehrere Publikationen (von mindestens fünf wissenschaftlichen Studien), die den Ein- und Ausschlusskriterien entsprechen.

B) Wenig Literatur: wenige oder keine Publikationen wissenschaftlicher Studienergebnisse, die den Ein- und Ausschlusskriterien entsprechen; möglicherweise Informationen aus Arbeiten verfügbar, die den methodischen Einschlusskriterien nicht entsprechen.

C) Keine wissenschaftlichen Studienergebnisse zur Fragestellung aufgefunden.

Wenig Anpassungsbedarf für die weitere Projektarbeit wird es in den Konstellationen A) und C) geben: bei ausreichend Literatur wird die Weiterarbeit protokollkonform fortgesetzt; werden keinerlei wissenschaftliche Informationen zur Fragestellung aufgefunden, schließt die Themenbearbeitung mit dieser Schlussfolgerung ab bzw. wird der Prozess der Feinabstimmung der Reviewfragestellung von vorne beginnen müssen.

Im Falle der Konstellation B) stehen die beiden oben genannten Wege offen - als dritte Möglichkeit können aber noch die Anforderungen an Evidenzbasis, den Ein- und Ausschlusskriterien für die Berücksichtigung von Studiendaten modifiziert werden. Diese Entscheidung ist grundsätzlich in der Arbeitsgruppe zu treffen, nicht durch den einzelnen Bearbeiter. Die Modifikationen werden begründet und protokolliert.

I C.3.2.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt

Um die zentrale Frage unseres HTA-Berichtes „Wie valide lässt sich das Frakturrisiko mit Hilfe der quantitativen Ultraschall-Verfahren im Vergleich zum DXA-Verfahren abschätzen?“ beantworten zu können, entwickelten wir zunächst eine Suchstrategie, in der wir die zu berücksichtigenden Einzelkomponenten spezifizierten.

Nach den Kriterien der Evidenzbasierten Medizin enthalten gut formulierte Fragen wenigstens vier Elemente, um relevante und präzise Antworten zu erhalten:

1. eine Beschreibung der interessierenden Patientencharakteristika,
2. die Darstellung der „Intervention“, bei der es sich je nach Art der Fragestellung oder des klinischen Problems um ein therapeutisches oder diagnostisches Verfahren, eine Ursache oder einen prognostischen Faktor handeln kann,
3. gegebenenfalls die Nennung einer Vergleichsintervention und
4. die Beschreibung der in Frage kommenden klinischen Zielgrößen.

Festlegungen für das exemplarische Kurz-HTA:

- Personen-/Patienten: primär keine Spezifikation der Population
- Intervention/Exposition: Ergebnisse der Quantitativen Ultraschalluntersuchung (QUS) des Skeletts
- Zielgrößen: Frakturen, (Osteoporose)
- Bereich: Risikoabschätzung, Prognose

Damit wurde festgeschrieben, welcher Typ von Primärstudie die adäquate Information liefert, nämlich die *prospektive Kohortenstudie*. Dies erlaubt nun die Formulierung von Ein- und Ausschlusskriterien sowie die Ableitung von adäquaten Suchbegriffen für die Datenbankrecherchen. Unsere Planung der Literaturrecherche nach hierarchischem Vorgehen wurde so festgelegt, dass zunächst nur nach Publikationen, die systematische Informationszusammenfassungen enthalten (z. B. HTA-Berichte, Systematische Literaturübersichten), gesucht wird. Diese Suche wurde in der Cochrane Library, in den HTA-Datenbanken (NHS CRD, DARE, NEED und HTA, HSTAT, INAHTA) sowie diversen Leitliniendatenbanken und den Publikationslisten der weltweit tätigen HTA-Institutionen durchgeführt (s. Dokumentation im Anhang).

- | | | |
|-------------|----|------------------------------------|
| Hierarchie: | 1a | HTA-Berichte |
| | 1b | Systematische Literaturübersichten |
| | 2 | Primärstudien |

Bei der Literaturrecherche zu dem Thema „Die Rolle der quantitativen Ultraschallverfahren zur Ermittlung des Risikos für osteoporotische Frakturen“ wurde zunächst in

der Cochrane Library, der INAHTA-Projektdatenbank sowie in Healthstar nach aktuellen HTA-Berichten (≤ 5 Jahre alt) gesucht.

Es fand sich eine Publikation (Homik & Hailey 1998) zum Thema, die eine Übersicht über die publizierte Literatur bis einschließlich 1997 bot. Durch das Auffinden dieses, den Einschlusskriterien entsprechenden, HTA-Berichtes konnten wir den Recherchezeitraum auf das Jahr 1998 bis zum Erstellungszeitpunkt unseres Kurz-HTA reduzieren, was eine erhebliche Zeitersparnis bedeutete.

Die Suche wurde auf systematische Übersichtsarbeiten ausgedehnt und schlussendlich Primärstudien recherchiert (Publikationszeitraum 1/1998 bis 12/2000).

Unter Berücksichtigung dieser Vorgehensweise war auch eine individuelle Zusammenstellung der Datenbankauswahl erforderlich. Im einzelnen wurde in folgenden Datenbanken recherchiert:

Cochrane Library	systematische Reviews, kontrollierte Studien, HTA-Berichte, ökonomische Bewertungen
INAHTA-Projektdatenbank	HTA-Berichte
HEALTHSTAR	HTA-Berichte, Reviews, Primärstudien
MEDLINE	Reviews, Primärstudien

Die Suchbegriffsauswahl erfolgte nach der orientierenden Literaturrecherche anhand vorliegender Zitate. Zunächst wurden die Schlüsselworte für den Bereich 'Zielkondition' (z. B. Osteoporose, Fraktur) festgelegt und sowohl mit Schlagworten (Mesh-Begriffen = Medical Subject Headings) und im Volltextmodus in den angegebenen Datenbanken gesucht. Anschließend wurde per Volltextsuche die Verfahrensseite (z. B. Ultraschalluntersuchung) recherchiert. Hierbei konnten keine Meshbegriffe verwendet werden, weil zu diesem Zeitpunkt für das neue Verfahren noch keine existierten. Die einzelnen Suchbegriffe wurden mit ODER verknüpft, beide Stränge abschließend mit UND aneinandergelunden.

Das detaillierte Vorgehen und die Ergebnisse dieser Suchstrategie sind im Anhang zu finden.

Die Zitate wurden nach zentral und peripher relevanter Literatur sortiert, wobei prospektive Studien zu Frakturrisiko und Ultraschallmessungen als primär relevant und Querschnitts- sowie Vergleichsstudien zu Ultraschallmessungen bei Kranken / Gesunden als sekundär relevant eingestuft wurden.

Ausgeschlossen wurden von vornherein Fallberichte, Tierstudien, sekundäre Osteoporosen, thematisch abweichende Zitate, Briefe und Editorials. Außerdem erfolgte eine Einschränkung der Publikationssprache auf Englisch, Deutsch, Französisch und Spanisch.

Ergebnisse der Literaturrecherche:

1. Cochrane Library

Tabelle 4: Ergebnisse der Literaturrecherche in der Cochrane Library

Die Literatursuche in der COCHRANE LIBRARY mit den Datenbanken (Issue 4, 2000):	
Cochrane Database of Systematic Reviews	15 Treffer
Database of Reviews of Effectiveness (DARE)	7 Treffer
INAHTA Projektdatenbank	3 Treffer
Cochrane controlled Trials Register (CENTRAL/CCTR)	80 Treffer
Economic evaluation database (NHS EED)	1 Treffer

Die Recherche der Cochrane Library ergab insgesamt **106** Treffer, die Selektion der Ergebnisse erfolgte nach den bekannten Ein- und Ausschlusskriterien anhand der nach Doublettencheck verbliebenen Abstracts. Die Protokolle (4 Treffer) wurden von vornherein ausgeschlossen.

- Cochrane Database of Systematic Reviews: Kein systematischer Review zur Fragestellung, alle 15 Treffer themenfremd.
- Database of Reviews of Effectiveness (DARE): Vier Literaturstellen, die die Einschlusskriterien erfüllen. Hierunter fällt auch der im Jahr 1998 erstellte HTA-Report 11 von Joanne Homik und David Hailey „Quantitative ultrasound for bone density measurement“.
- INAHTA Projektdatenbank: Drei Treffer, auch hier findet sich das Abstract des schon identifizierten HTA-Berichtes, die anderen beiden Abstracts sind nicht themenrelevant.
- Cochrane controlled Trials Register (CENTRAL/CCTR): Unter den 80 Treffern im Cochrane Controlled Trials Register fanden sich 29 Literaturzitate, die die Einschlusskriterien erfüllten, davon verblieben nach Doubletten-Check noch neun, die allesamt Abstracts von Kongressen, Workshops, Konferenzen und Meetings darstellten und somit ausgeschlossen wurden.

Medline und Health Star

Die Ergebnisse dieser Recherchen reflektieren die Tatsache, dass zu einem frühen Zeitpunkt der Projektarbeit erwogen wurde, einen Abgleich zwischen umfassender und "gekürzter" Datenbankrecherche vorzunehmen. Nachdem sich herausstellte, dass dieser Ansatz aufgrund des erheblichen Zeitaufwandes nicht durchgehalten werden konnte, wurde für die Aufarbeitung der eigentlich umfassend konzipierten

Recherche nachträglich die im Schnellverfahren einzuführenden Beschränkungen umgesetzt.

Medline und Health Star (synchron recherchiert beim DIMDI): In MEDLINE (ab 1966) und HEALTH STAR (ab 1975) ergab die Recherche **1607** Treffer, nach Doublettencheck **915**. Die Titel der Zitate sowie die Abstracts wurden von zwei unabhängigen Untersuchern manuell nach den folgenden Ein- und Ausschlusskriterien auf potentielle Relevanz für das HTA-Vorhaben selektiert:

Tabelle 5: Relevanzkriterien für die Literaturselektion

Relevanz	Kriterien
zentral	Ultraschallmessung, Frakturrisiko, prospektive Studie
peripher	Ultraschall und DXA, Querschnittsstudie, Ultraschall, populationsbezogene Stichprobe: Normwerte, Frakturpatienten im Vergleich zu Personen ohne Frakturen
Ausschluss	Sekundäre Osteoporosen; in vitro Studien, Tierstudien, thematisch abweichend, Briefe, Editorials, Fallberichte; Sprache nicht Englisch, Deutsch, Französisch, Spanisch; eindeutig unsystematische Übersichtsartikel
unklar	anhand der Angaben im Abstract nicht eindeutig zuzuordnen

Das Ergebnis der ersten Bewertung der Literaturzitate ist wie folgt:

1. zentral relevant: **43** Treffer
2. peripher relevant **208** Treffer
3. unklar **27** Treffer
4. Ausschluss **637** Treffer

Der HTA-Report 11 aus dem Jahr 1998 wird einer Qualitätsbewertung unterzogen: er enthält eine systematische Zusammenfassung von Studien des adäquaten Studientyps, somit wurde die weitere Aufarbeitung der Ergebnisse der Medline-/Healthstar-Recherche auf die Jahrgänge 1998 – 2001 begrenzt. Alle Literaturstellen mit Publikationsdatum vor **1998** wurden zunächst zurückgestellt.

- Doublettencheck gesamt: 488 Publikationen
- 1. Selektion anhand der Abstracts: 126 Publikationen
- Bestellung

Die vollständigen Artikel der 126 Publikationen wurden beschafft und erneut durch zwei unabhängige Untersucher nach den gleichen Ein- und Ausschlusskriterien sor-

tiert. Die entstandenen Unstimmigkeiten wurden zwischen den Reviewern ausdiskutiert bis eine einstimmige Bewertung resultierte.

Von den 43 Reviews wird nur ein einziger als systematischer Review identifiziert. Die 83 verbliebenen Primärstudien werden auf Studiendesign und Thematik überprüft. Es erfolgt die Einteilung in (longitudinale) Prognosestudien und Diagnosestudien (Querschnittsstudien). Nach dem Einteilungsschema für diagnostische Studien nach Köberling (vergl. auch SOP 3) werden innerhalb der Gruppe von Diagnosestudien noch Phase-II-Studien, d. h. Ultraschall bei bekanntem Krankheitsstatus von den Phase-III-Studien ("reale" Studienpopulation) unterschieden.

Zweite Selektion anhand der vollständigen Artikel 39 Publikationen

Es verbleiben also nach erneuter Bewertung und Sichtung durch zwei unabhängige Untersucher :

- 7 Prognosestudien
- 30 Phase-II-Studien (diagnostisch)
- 1 Systematischer Review (Prins et al. 1997)
- 1 HTA-Bericht (Homik/Hailey 1998)

Insgesamt finden sich also nach der hierarchischen Auswahl noch **39** Literaturstellen (von ursprünglich insgesamt **916**), die nun nach eigens erstellten Checklisten / Stufenplänen bewertet werden (vergl. SOP 3).

I C.3.2.4 Diskussion: Bewertung, eigene Erfahrungen

Gelingt es, bereits in der Phase der orientierenden Literaturrecherche einen guten, passenden und themenrelevanten Review bzw. HTA-Bericht neueren Datums zu identifizieren, bedeutet dies eine erhebliche Zeitersparnis, da die Literaturrecherche auf den Zeitraum vom Erstellungsdatum des gefundenen Berichtes bis zum aktuellen Publikationstermin reduziert werden kann. Im hier berichteten Vorhaben bedeutete das frühzeitige Auffinden des HTA-Reports aus Kanada, dass nur noch Jahrgänge \geq 1998 zu berücksichtigen waren. Hätte man den HTA-Bericht nicht gefunden, hätte man noch 10 Jahrgänge mit weiteren 428 Publikationen bewerten müssen.

Allerdings muß angemerkt werden, dass die Suche insbesondere bei der Bewertung neuerer Technologien – wie in unserem Fall – auf den Zeitraum der letzten 3 - 5 Jahre beschränkt werden kann, da ältere Reviews sicherlich nicht mehr den aktuellen Stand einer Technologie widerspiegeln können. Bei der Eingrenzung des Publikationszeitraumes spielt auch eine entsprechende Einschätzung von Expertenseite eine große Rolle.

Ob und wie sich die Beschränkung der Recherche auf die Kerndatenbanken Cochrane Library, Medline und Healthstar unter Nichtberücksichtigung von vor allem der zu

Medline komplementären Datenbank Embase (Elsevier) auf die Ergebnisse und Schlussfolgerungen des Kurz-HTA auswirkt, wird in einem Folgeprojekt zur Validierung der Kurz-HTA-Methodik zu zeigen sein.

Es empfiehlt sich, die Ergebnisse der Datenbankrecherchen in einem elektronisch weiterverarbeitbaren Format zu speichern. Literaturverwaltungsprogramme wie z. B. Reference Manager® vereinfachen Handhabung und Verwalten der gesammelten Referenzen. Dabei sind vor allem die Möglichkeiten zusätzliche Information (z. B. aus der Bearbeitung) in eigens definierten Feldern abzuspeichern und die automatisierbare Erstellung von Referenz- und Publikationslisten nach standardisierten Layout-Vorgaben hilfreiche Werkzeuge, die Zeiteinsparungen vor allem bei den vielfältigen Dokumentationspflichten ermöglichen.

Die Selektion der Literaturzitate erfolgte durch zwei unabhängige Untersucher. Dies halten wir für unabdingbar, da auf diese Weise Einseitigkeiten entdeckt und ausdiskutiert werden konnten. Es ist auf jeden Fall zweckmäßig und zeitsparend, nicht alle bei der Recherche identifizierten Publikationen im Volltext zu beschaffen (126 Publikationen) und zu bearbeiten, sondern bereits auf der Basis der Abstracts die offensichtlich auszuschliessenden Studien auszusortieren. Wir halten dieses Prozedere für sinnvoll, da unserer Meinung nach die Gefahr, wichtige Informationen zu übersehen, sehr gering ist. Man kann davon ausgehen, dass die Verfasser von Abstracts die wesentlichen Merkmale ihrer Publikationen inhaltlich vollständig zusammenfassen und entsprechend verschlagworten.

Eine Beschleunigung dieses Arbeitsschrittes erfolgt also durch die Vorauswahl auf der Basis von Abstracts (und nicht von Vollartikeln), durch die standardisierte Anwendung von Ausschlusskriterien, indem case reports, animal studies, Editorials, Leserbriefe, Kommentare, Konferenzen etc. ausgeschlossen werden sowie durch Einschränkung der Publikationssprache.

Die vollständigen Artikel der Restmenge wurden per Online-Bestellung beschafft; dies beschleunigt die Erstellung des HTA-Berichtes bei allerdings erhöhtem Mitteleaufwand (4 € / Kopie bei Lieferung innerhalb von drei Werktagen anstelle 0,56 € / Kopie bei Lieferung nach frühestens drei Wochen).

Ergänzende Literaturrecherchen z. B. die handdurchsuchte Literatur, nicht elektronisch erfasste Kongressberichte oder Kontakt zu Verfassern und Studienleitern wurden im Rahmen unseres Kurz-HTA nicht durchgeführt, da sie ausgesprochen zeitintensiv sind. Inwieweit dies zu Verzerrungen der Ergebnisse führen kann, ist auch in der Literatur nicht hinreichend geklärt.

Die von uns in der ersten Projektphase durchgeführte Medline-Recherche mit der nachfolgend beschriebenen Suchstrategie wurde zur Prüfung an Frau Sigrid Droste, MHH Hannover, German Scientific Working Group Technology Assessment for Health Care geschickt. Die Verbesserungsvorschläge, Detailspekte betreffend, wurden entsprechend eingearbeitet.

Auch die Liste der ein- und ausgeschlossenen Publikationen wurden zur Überprüfung an die Arbeitsgruppe Hannover gegeben. Wir bedanken uns an dieser Stelle für die unkomplizierte Zusammenarbeit.

I C.3.2.5 Schlussfolgerung: SOP 2

Vorschlag für eine generalisierbare Arbeitsanleitung

1. Spezifikation der zentralen Berichtsfragestellung nach:

- Charakteristika der Patienten- / Personenpopulation
- Charakteristika der Krankheit / des Gesundheitsproblems
- Intervention / Exposition (ggf. Alternativen)
- Zielgrößen
- Zuordnung der Fragestellung zu einem Bereich medizinischer Entscheidungen (z. B. Indikation zum Einsatz von Diagnostik oder Therapie; Risikoabschätzung, Prognosestellung). Hierdurch wird bestimmt, welches Primärstudiendesign die adäquaten Informationen liefert.

Die konsequente Ausformulierung an diesen Stellen ermöglicht:

- die Formulierung von Ein- und Ausschlusskriterien (Zieldefinition für die Recherche) sowie
- die Ableitung von inhaltlichen und methodischen Suchbegriffen für die Datenbanksuchen.

2. Planung der Literaturrecherche nach hierarchischem Vorgehen:

- Publikationen, die systematische Informationszusammenfassungen enthalten: HTA-Berichte, systematische Literaturübersichten
- Primärstudien:
 - zur Abdeckung der „Lücke“ zwischen Redaktionsschluss (der Übersichten) und Gegenwart
 - wenn keine systematischen Übersichten existieren

3. Datenbankauswahl

- HTA-Berichte: CRD-HTA, INAHTA, ISTAHC-Datenbank komplett
- Systematische Übersichten: CDSR, DARE: komplett (Medline: unter Verwendung einer spezifischen Suchstrategie für systematische Reviews)
- Primärstudien: Medline; Auswahl der Jahrgänge s. unter Punkt 2
- Weitere Datenbanken: nur in Ausnahmefällen

4. Such- und Selektionsstrategie:

- Sensitiv unter inhaltlichen Gesichtspunkten; anschließend manuelle Selektion der Abstracts nach Ein- und Ausschlusskriterien

5. Beschaffung:

- Den Ein- und Ausschlusskriterien entsprechend; bis zum Vorliegen der Ergebnisse einer Validierungsstudie: bei unklaren Angaben im Abstract oder fehlendem Abstract sollte ebenfalls der Volltext bestellt werden. Die Nutzung von Lieferdiensten wie Subito[®] ermöglicht - allerdings bei höherem Ressourcenaufwand - eine deutliche Abkürzung der Lieferfristen im Vergleich zum regulären Fernleihverkehr im deutschen Bibliotheksverbund.

6. Bestandskonferenz:

- Entscheidet, ob angesichts der Menge und der (voraussichtlichen) Qualität der verfügbaren Literatur protokollkonform weitergearbeitet, die Ein- und Ausschlusskriterien modifiziert oder das Projekt abgebrochen werden muss.

I C.3.3 SOP 3: Vorgehen bei der kritischen Informationsbewertung; Bewertungsinstrumente

I C.3.3.1 Hintergrund: Status Quo (Ausführliches HTA)

Die Vorbereitung jeglicher Art von Informationssynthese umfasst die kritische Bewertung der Primärstudien. Die beiden Aspekte, die von der Bewertung erfasst werden müssen, sind zum einen die methodische Qualität (**interne Validität**), zum anderen die Relevanz der Materialien für die Fragestellung (**externe Validität**).

Interne Validität

Interne Validität drückt aus, ob die Resultate einer Studie für die untersuchte Stichprobe von Patienten Gültigkeit haben. Sie ist „intern“, weil sie für die Bedingungen der betreffenden Patientengruppe, die beobachtet wurde, gilt und nicht unbedingt für andere. Die interne Validität klinisch-epidemiologischer Untersuchungen wird dadurch bestimmt, wie gut die Planung, die Datensammlung und die Analysen durchgeführt werden - d. h. wie gut das Studiendesign das Auftreten von systematischen Fehlern (Biases) kontrolliert und damit für unverzerrte Ergebnisse sorgt. Die interne Validität ist eine notwendige Voraussetzung für die Verwertbarkeit von Studienergebnissen.

Einordnung nach Evidenzhierarchien

Um sicher zu stellen, dass nur Primärstudien mit hoher interner Validität zum Einschluss in die Literaturübersicht ausgewählt werden, wird ein zweischrittiges Vorgehen eingesetzt. Der erste Schritt, der im Prinzip schon im Rahmen der in SOP 2 beschriebenen Literaturselektion vorgenommen wird, sortiert die in der Literaturrecherche als relevant identifizierten publizierten Studien nach methodischen Qualitätsas-

pekten. Hierzu werden die verschiedenen Studien in eine sogenannte „Hierarchie der Evidenz“ eingeordnet, wobei Studientypen, die weniger anfällig für systematische Fehler sind, höher bewertet werden als solche, bei denen systematische Fehler nicht ausgeschlossen werden können. Tabelle 6 zeigt ein typisches Beispiel einer „Evidenzhierarchie“ für Fragestellungen aus dem Bereich Wirksamkeit von Interventionen (auch für Argumentationen aus dem Gebiet der unerwünschten Wirkungen von Interventionen liefern die hier angeführten Studiendesigns prinzipiell die validesten Ergebnisse – allerdings sind sie aus ethischen Gründen in der Regel nicht durchführbar). Da diese Evidenzhierarchie von einer Arbeitsgruppe entworfen wurde, die selber mit der Verfassung von systematischen Reviews befasst ist, findet sich ein Einzelstudiendesign an der Spitze der Hierarchie. Überarbeitungen der „Hierarchie“ für Therapiestudien durch Institute, die überwiegend mit der Umsetzung von wissenschaftlichen Informationen befasst sind (z. B. CEBM 2001), führen als Klasse-I-Evidenz zumeist eine Meta-Analyse von Ergebnissen methodisch einwandfreier randomisierter kontrollierter Studien an.

Tabelle 6: Evidenzhierarchie für Therapiestudien nach CRD (The University of York, NHS Centre for Reviews and Dissemination 1996)

Stufe	Design
I	Methodisch einwandfreie randomisierte, kontrollierte Studie
II-1a	Methodisch einwandfreie kontrollierte Studie mit Pseudorandomisierung
II-1b	Methodisch einwandfreie kontrollierte Studie ohne Randomisierung
II-2a	Methodisch einwandfreie Kohortenstudie (prospektiv) mit aktuellen Kontrollen
II-2b	Methodisch einwandfreie Kohortenstudie (prospektiv) mit historischen Kontrollen
II-2c	Methodisch einwandfreie Kohortenstudie (retrospektiv) mit aktuellen Kontrollen
II-3	Methodisch einwandfreie Fall-Kontrollstudie (retrospektiv)
III	Große Unterschiede im Vergleich zwischen verschiedenen Orten und/oder Zeiten innerhalb derselben Population, mit oder ohne Intervention
IV	Meinungen von ausgewiesenen Experten, beschreibende Studien, Berichte von Expertengremien

Für eine evidenz-basierte Argumentation in Leitlinien oder HTA-Berichten ist es häufig aber auch erforderlich, Informationen zu den Auswirkungen einer „Exposition“ auf das Krankheitsrisiko oder -outcome (Prognose) zu berichten. Unter Exposition sind in diesem Sinne keine Interventionen zu verstehen sondern Faktoren, die – häufig in unterschiedlicher Ausprägung - vorhanden sind oder nicht (z. B. Zigarettenrauchen, Bewegungsmangel, erhöhte Cholesterinwerte u. ä.). Ihre Auswirkungen müssen in observationellen (beobachtenden) Studiendesigns untersucht werden, da die „Exposition“ nicht von der Studienleitung verabreicht werden darf. Die Untersuchungs- bzw. Auswertungsgruppen werden vielmehr in Abhängigkeit von ihrem Expositionsstatus zusammengestellt. Designs, die zur Untersuchung der Zusammenhänge von Exposition und Outcome herangezogen werden, sind Kohortenstudien, Fall-

Kontrollstudien aber auch Querschnittsuntersuchungen. Auch diese Studien sind (unterschiedlich) anfällig für systematische Verzerrungen.

Die Arbeitsgruppe des Center for Evidence-based Medicine in Oxford hat auch für prognostische Fragestellungen eine Hierarchie der Evidenz zusammengestellt (Tabelle 7).

Diese Levels können noch Abstriche in Form eines Minus (–) erhalten, wenn sie keine schlüssige Antwort auf die Ausgangsfragestellung bieten. Dies kann der Fall sein wenn:

- das Konfidenzintervall um das Ergebnis von Einzelstudien so weit ist, dass, bei fehlender statistischer Signifikanz, der klinisch relevante Endpunkt mit eingeschlossen ist,

oder

- die eingeschlossenen Studien in einer Literaturübersicht erhebliche Heterogenität aufweisen.

Tabelle 7: Evidenzhierarchie für Prognosefragestellungen (CEBM 2001; Ü.d.A.)

Stufe	Design
1a	Systematische Übersicht über homogene Inzeptionskohortenstudien; in unterschiedlichen Populationen validierte Entscheidungsregeln.
1b	Einzelne Inzeptionskohortenstudie mit einem Follow-Up > 80% ; in einer Population validierte Entscheidungsregel.
1c	Alles-oder-Nichts Fallserie*
2a	Systematische Übersicht über homogene retrospektive Kohortenstudien oder unbehandelte Kontrollarme von RCTs.
2b	Retrospektive Kohortenstudien oder Ergebnisse des unbehandelten Kontrollarmes aus einem RCT. Ableitung von Entscheidungsregeln aus Subpopulationen.
2c	"Outcome" Studien
3a / 3b	<i>(entfällt bei Prognosestudien)</i>
4	Fallserien und Kohortenstudien schlechter methodischer Qualität
5	Expertenmeinung ohne die Möglichkeit zu einer kritischen Überprüfung bzw. Informationen aus Grundlagen- (Laborforschung), Ableitungen aus physiologischen Zusammenhängen oder "first principles"

* erfüllt wenn (fast) alle Personen unter Exposition ein bestimmtes Outcome erleiden - oder ohne Exposition (fast) keiner (Beispiel: Phokomelien nach Conterganexposition der Mütter in der Schwangerschaft).

Für Fragestellungen aus dem Gebiet „Evaluation von diagnostischen Verfahren“ stellte sich die Entwicklung einer Evidenzhierarchie als deutlich schwieriger heraus. Es schien sinnvoll, diagnostische Studien, in Analogie zu den Pharmatherapiestudien, den Phasen I bis IV zuzuordnen. Köbberling et al. präsentierten 1989 hierzu

einen Entwurf, der zumindest im deutschsprachigen Raum Akzeptanz gefunden hat (vergl. Tabelle 8).

Tabelle 8: Phaseneinteilung der Diagnoseevaluierung (nach Köbberling et al. 1989)

Stufe	Design
Phase 1:	Technische und methodische Voruntersuchungen als Basis für die weitere klinische Evaluierung Analytische Sensitivität und Spezifität bei bekanntem Krankheitsstatus Reproduzierbarkeit, Richtigkeit, Akzeptanz, Risiko
Phase 2:	Verteilung der Testergebnisse bei Gesunden und Kranken Schätzung der diagnostischen Sensitivität und Spezifität bei bekanntem Krankheitsstatus
Phase 3:	Die sogenannte „kontrollierte diagnostische Studie“, durchgeführt in der konkreten klinischen Anwendungssituation unverzerrte Schätzung von diagnostischer Sensitivität und Spezifität Schätzung der Prävalenz und der prädiktiven Werte
Phase 4:	„Wirksamkeitsuntersuchung“ der Anwendung (RCT)

Vor dem Hintergrund dieser Einteilung wird klar, dass zur Beurteilung der „Wirksamkeit“ (seine Fähigkeit in der konkreten klinischen Einsatzsituation, Erkrankte und Nichterkrankte zuverlässig zu diskriminieren) eines diagnostischen Testverfahrens mindestens Studien der Phase III erforderlich sind. Im Gegensatz zu den Therapiestudien existiert allerdings für diagnostische Studien noch kein allgemein akzeptierter Konsens, welches Design als die höchste Evidenzstufe gelten sollte. In der Evidenzhierarchie für diagnostische Studien des Centre for Evidence Based Medicine in Oxford ist als höchste Evidenzklasse für diesen Typ Fragestellung die Meta-Analyse von homogenen Kohortenstudien angeführt (CEBM 2001), während die Richtlinien des Bundesausschusses der Ärzte und Krankenkassen (BUB-Richtlinien, Richtlinien über die Bewertung ärztlicher Untersuchungs- und Behandlungsmethoden gemäß § 135 Abs. 1 SGBV, Bundesanzeiger Nr. 56 am 21.3.2000) hier die Ergebnisse einer randomisierten Wirksamkeitsuntersuchung (Phase IV Studie nach Köbberling) fordern. Ein sehr komplexes Modell zur Bewertung von Diagnosestudien wurde von Flynn (1996) publiziert, basierend auf den Vorschlägen von Kent und Larson (1988). Hier wird zunächst die Information, die der Verfahrensbewertung entnommen werden soll, einer von sechs aufeinander aufbauenden Dimensionen zugeordnet (technical efficacy, diagnostic accuracy efficacy, diagnostic thinking efficacy, therapeutic efficacy, patient outcomes efficacy und societal efficacy). Für jede Dimension sind die optimalen Studiendesigns festgelegt (z. B. RCTs für die obersten drei Dimensionen), welche ihrerseits auf interne und externe Validität zu überprüfen sind.

Qualitätsprüfung im Detail

Da das Studiendesign allein nicht genügend über die Qualität der Untersuchung aussagt, ist es unumgänglich in einem zweiten Schritt, nach der Zuordnung zu einer Evi-

denzstufe, eine genauere Bewertung der einzelnen Studien auf ihre methodische Validität hin vorzunehmen.

Zentrale Frage bei Primärstudien ist immer (unabhängig von der Review-Fragestellung): ist die Studie so konstruiert, dass die Ergebnisse die Realität im interessierenden Kontext abbilden? Eine Verzerrung der Ergebnisse bzw. Schlussfolgerungen durch zufällige oder systematische Fehler sowie Confounding (Verwechslung) soll ausgeschlossen sein.

Hierzu ist eine Vielzahl unterschiedlich umfangreicher Bewertungsschemata und Checklisten entwickelt worden (Übersicht für Therapiestudien z. B. bei Moher et al. 1995). In der Cochrane Collaboration verwendete Checklisten hinterfragen zu diesem Zweck vier bekannte Ansatzstellen für systematische Verzerrungen von Studienergebnissen bei randomisierten kontrollierten Studien (RCTs):

- Systematische Fehler bei der Zuordnung zu Interventions- bzw. Kontrollgruppen (**selection bias**)
- Systematische Fehler bei der Behandlung der Gruppen, abgesehen von der zu untersuchenden Intervention (**performance bias**)
- Systematische Fehler durch Studienabbrecher (**exclusion bias**)
- Systematische Fehler bei der Erfassung / Beobachtung von Outcomes (**detection bias**)

Diese Ansatzstellen für systematische Fehler sind natürlich nicht nur in randomisierten kontrollierten Studien sondern prinzipiell in allen klinischen Studiendesigns überprüfbar. Abhängig vom Design wird den verschiedenen Aspekten allerdings unterschiedlich viel Aufmerksamkeit gewidmet. Hauptfehlerquelle (und damit Ursache für verzerrte Ergebnisse) in RCTs ist die nicht-zufällige Zuordnung der Studienteilnehmer zu Verum- und Kontrollgruppen. Eine Qualitätscheckliste für RCTs wird daher diesem Aspekt viel Aufmerksamkeit widmen. Verzerrte Ergebnisse in Kohortenstudien dagegen entstehen viel häufiger durch den Verlust eines großen Anteils von Studienteilnehmern in diesen oft auf mehrere Jahre angelegten Projekten. Eine Checkliste für Kohortenstudien wird daher dem Aspekt „Follow-Up“ mehr Aufmerksamkeit widmen. Auf diese Art lassen sich für alle Studiendesigns unterschiedliche Aspekte identifizieren, die im Rahmen einer Qualitätsprüfung abzufragen sind.

Bei der Beurteilung der methodischen Qualität von „Sekundärpublikationen“, d. h. solchen Publikationen, die bereits Informationssynthesen enthalten (HTA-Berichte, Reviews und Meta-Analysen), ist die Frage zu beantworten, ob in der Übersichtsarbeit der wissenschaftliche Kenntnisstand unverzerrt abgebildet und beurteilt wird. Die Checklisten für Sekundärpublikationen berücksichtigen das systematische Vorgehen bei Literaturrecherche, -selektion, Qualitätsbewertung und Informationssynthese sowie Konsequenz von Diskussion und Schlussfolgerungen. In der Checkliste der deutschen HTA-Arbeitsgruppe für Kontextdokumente (HTA-Berichte, Leitlinien) wird

in einem zweiten Teil konkret nach der Umsetzbarkeit der Schlussfolgerungen im Kontext des bundesdeutschen Gesundheitssystems gefragt. Es sollen Unterschiede zu in Deutschland gegebenen Verhältnissen (Dimensionen: Epidemiologie der Zielkondition, Versorgungskontexte, -bedingungen und –prozesse, Vergütungssysteme, Kosten, Preise, Indikationsstellung, Patienten- und Providerpräferenzen) aufgezeigt werden, die eine Übertragbarkeit der Ergebnisse in Frage stellen oder verhindern.

Die Beurteilung der methodischen Studienqualität mit Hilfe von standardisierten Bewertungsinstrumenten hat den Vorteil, dass sie den Prüfungsvorgang vereinheitlicht, besser erlernbar macht und eine Vergleichbarkeit der Ergebnisse gewährleistet.

Drei Typen von Bewertungsinstrumenten werden unterschieden: Komponentenlisten, Checklisten und Scoresysteme. In Komponentenlisten wird lediglich das Vorhandensein einiger weniger Komponenten, die für das Erreichen von unverzerrten Studienergebnissen für essentiell gehalten werden, überprüft (z. B. bei RCTs: geheime und unbeeinflussbare Zuordnung der Teilnehmer zu Verum- und Kontrollgruppen, Verblindung, Intention-to-treat Analyse).

In Checklisten wird eine größere Anzahl qualitätsrelevanter Items abgefragt, ohne dass ihnen Punktwerte zugeordnet werden. Scoresysteme bestehen aus Checklisten, in welchen den einzelnen Items und Abstufungen Punktwerte zugeordnet werden. Die Punktwerte können für alle Items gleich sein oder aber gestaffelt, je nach der Bedeutung, die den Items zugemessen wird. Mithilfe dieser Instrumente lassen sich „Qualitätspunkte“ für einzelne Studien berechnen. Bewertungen von Studien, die mit Hilfe von Scoresystemen vorgenommen wurden, wirken auf den ersten Blick differenzierter als solche, die durch einfaches Abfragen von Komponenten erhalten wurden.

Allerdings ist es zweifelhaft, ob sehr komplexe Bewertungen tatsächlich einen Informationsgewinn bringen – die große Zahl der abgefragten Items gibt häufig nur Auskunft über die Qualität der Studienpublikation und nicht der Studie selber (Ziegler et al. 2001). Sie sind einerseits zeitaufwendig und schaffen andererseits eine Vielzahl von Qualitätskategorien, die eine Informationssynthese erschweren können. In den meisten Reviews bestimmt ein dichotomes Ergebnis der Qualitätsbewertung, ob die Ergebnisse der bewerteten Studie in die systematische Literaturübersicht einfließen oder nicht. Einige Reviewautoren verwenden abgestufte Bewertungssysteme und haben damit die Möglichkeit differenzierter zu bestimmen, wie groß der Einfluss eines Einzelstudienresultates auf die Gesamtaussage der Literaturübersicht ist. Die graduierte Bewertung liefert die Basis für unterschiedliche Interpretationsmöglichkeiten, ermöglicht Sensitivitätsanalysen auf der Grundlage methodischer Studienqualität oder kann als Gewichtungsfaktor in statistischen Analysen (Meta-Analysen) dienen.

Arbeiten wie die von Juni et al. (1999) weisen aber darauf hin, dass allein die Konstruktion von komplexen Scoreinstrumenten so viel Subjektivität in die Studienbewertung hineinträgt (durch unterschiedliche Schwerpunktsetzungen), dass die Verwendung von unterschiedlichen Qualitätsbewertungsinstrumenten zu dramatisch unterschiedlichen Ergebnissen der Meta-Analyse und damit Schlussfolgerungen für die

Praxis führt. Ein konsistenter Zusammenhang zwischen Punktescore und berichteter Effektgröße konnte aber für keines der 25 von ihnen getesteten Instrumente nachgewiesen werden, wohl dagegen für einzelne Aspekte des Studiendesigns. Die Autoren empfehlen daher, auf Scores zu verzichten und die Studienbewertung mit einfachen Komponentenlisten vorzunehmen.

Externe Validität

Vorbedingung zur Beurteilung der externen Validität ist die exakte Definition der zu beantwortenden Forschungsfragen. Konnte die Literaturrecherche noch unter relativ breiter thematischer Eingrenzung erfolgen, ist es für die Erstellung von Bewertungskriterien mindestens erforderlich festzuhalten:

- die Intervention / Technologie von Interesse,
- die system- und indikationsbedingte Anwendungssituation sowie die betroffenen Patienten- bzw. Klientengruppen,
- die erwünschten Outcomeparameter.

Einflüsse, die an dieser Stelle Verzerrungen der gewonnenen Information bewirken können, sind die einseitig beeinflusste Auswahl der in die Auswertung einbezogenen Primärmaterialien (s. SOP 2). Im Gegensatz zur Beurteilung der internen Validität erfordert die Beurteilung der externen Validität vor allem klinischen und / oder verfahrensbezogenen Sachverstand - geht es doch um die Beurteilung der Übertragbarkeit von Studien- bzw. Reviewergebnissen unter den oben genannten Aspekten auf die spezifische eigene Entscheidungssituation.

Checklisten der German Scientific Working Group for HTA

Im Anhang finden sich die von der German Scientific Working Group entwickelten Checklisten, die in den bisherigen HTA-Gutachten zur Bewertung und Dokumentation der Studienqualität verwendet wurden. Die Bewertung wurde grundsätzlich durch zwei Autoren vorgenommen, die ausgefüllten Checklisten wurden als Hintergrunddokumente in die Datenbank eingebunden.

- Qualitätsbeurteilung von Kontextdokumenten (HTA-Berichte, Leitlinien)
- Qualitätsbeurteilung von Systematischen Reviews (ggf. mit Meta-Analysen)
- Qualitätsbeurteilung von Primärstudien (ausgenommen: diagnostische Studien)
- Qualitätsbeurteilung von diagnostischen Studien

Checkliste 1a der German Scientific Working Group zur Bewertung der methodischen Qualität und der Relevanz der dargestellten Informationen: Das Instrument wurde konzipiert zur Bewertung von HTA-Berichten, Leitlinien und anderen Dokumenten, in denen Informationszusammenfassungen in einem entscheidungsrelevanten „Kontext“ eingebunden sind. In insgesamt 40 Items werden in sechs Dimensio-

nen Qualitätskriterien erfasst. Unter den Items wurden zwei Typen unterschieden: solche, die methodische Qualitätskriterien abfragen (Q) und solche, die zur Dokumentation von Kontextinformationen auffordern (I). Die qualitätsrelevanten Items werden nach ihrer Wichtigkeit weiter bezeichnet mit A, B oder C. Mit QA werden Items gekennzeichnet, die zentral wichtige Qualitätskriterien ansprechen, danach folgen die mit QB und QC bezeichneten Fragen. Für alle Fragen gibt es die Antwortmöglichkeiten „Ja“, „Nein“ und „unklar“. Das Ergebnis der Qualitätsbewertung wird dichotom angegeben: Die in der Publikation berichteten Ergebnisse werden in den Review eingeschlossen – oder nicht. Um für einen Einschluss in Frage zu kommen, müssen alle mit QA gekennzeichneten Fragen mit „Ja“ beantwortet werden. Die Verteilung der Items auf die Dimensionen ist in der Checkliste 1a wie folgt:

- Fragestellung und Kontext (14 Items, 1 QA, 13 I)
- Methodik der Informationsgewinnung (4 Items, 1 QA, 3 QB)
- Methodik der Bewertung und Dokumentation (5 Items, 1 QA, 4 QC)
- Methodik der Informationssynthese (3 Items, 3 I)
- Ergebnisse und Schlussfolgerungen (6 Items, 1 QA, 1 QB, 1 QC, 3 I)
- Übertragbarkeit der internationalen Ergebnisse / Schlussfolgerungen (7 Items, alle I).

Checkliste 1b dient der Bewertung von systematischen Reviews mit / ohne Meta-Analyse, die als alleinstehende wissenschaftliche Arbeit konzipiert wurden ohne in einen spezifischen entscheidungsrelevanten Kontext eingebunden zu sein. Insgesamt 22 Items in fünf Dimensionen fragen Merkmale ab, die Aufschluss über die interne und externe Validität des Reviews geben sollen. Die Fragen verteilen sich wie folgt:

- Relevanz der Fragesstellung (1 Item, 1 QA)
- Informationsgewinnung (4 Items, 1 QA, 3 QB)
- Informationsbewertung (5 Items, 1 QA, 1 QB, 3 QC)
- Informationssynthesen (5 Items, 2 QA, 2 QB, 1 QC)
- Schlussfolgerungen (7 Items, 1 QA, 2 QB, 4 I)

Checkliste 2a dient der Bewertung von allen Primärstudien, mit Ausnahme von Studien, die diagnostische Verfahren evaluieren. Das Instrument fordert zunächst die Identifikation und Spezifikation des Studiendesigns und erfasst dann insgesamt 30 Items in den sieben Dimensionen:

- Auswahl der Studienpopulation (6 Items, 4 QA, 2QB)

- Zuordnung und Studienteilnahme (5 Items, 3 QA, 1 QB, 1 QC)
- Intervention / Exposition (5 Items, 3 QA, 2 QB)
- Studienadministration (2 Items, beide QB)
- Outcome Messung (4 Items, 1 QA, 1B, 1 QC, 1 I)
- Drop Outs (5 Items, 3 QA, 2 QB)
- Statistische Analyse (3 Items, 1 QA, 1 QB, 1 I)

Auch dieses Instrument dichotomisiert das Bewertungsergebnis: Studien, für die alle QA Items mit „Ja“ beantwortet werden, werden in den Review eingeschlossen.

Checkliste 2b wird zur Beurteilung von Studien herangezogen, deren Ziel die Evaluation von diagnostischen Testverfahren ist. Insgesamt werden 14 Items in den folgenden vier Dimensionen abgefragt:

- Beschreibung der Ausgangssituation (4 Items, alle 4 QA)
- Durchführung der Prüfung (4 Items, 2 QA, 2 QB)
- Ergebnispräsentation (2 Items, 1 QA, 1 QB)
- Diskussion (4 Items, alle 4 QB)

[Da aus Zeitgründen eine gesundheitsökonomische Bewertung im Rahmen des vorliegenden Kurz-HTAs nicht durchgeführt wurde, wird an dieser Stelle auch auf eine ausführliche Darstellung der Auswahl- und Beurteilungskriterien der methodischen Qualität gesundheitsökonomischer Studien verzichtet. Der Status Quo für ausführliche HTA-Berichte „Checkliste zur Beurteilung der methodischen Qualität gesundheitsökonomischer Studien“ ist jedoch im "Toolkit" nachzulesen].

Auswahl- und Beurteilungskriterien dienen also der Identifikation von inhaltlich relevanten, qualitativ-methodisch hochwertigen Primärmaterialien für die Verfahrensbeurteilung. Die inhaltliche Relevanz wird über die Definition von themenspezifischen präzisen Einschlusskriterien gewährleistet, für die methodische Qualität wurde ein Vorgehen nach Checklisten (und Empfehlung der Anlage eines Datasheets) vereinbart, da dieses Verfahren gegen subjektive Einflüsse weitgehend unempfindlich ist und gleichzeitig gute Dokumentationsmöglichkeiten bietet.

I C.3.3.2 Methoden: Modifikations-/Kürzungspotential

Die bisherige Arbeit an HTA-Projekten hat zwei für die Qualitätsbewertung relevante Erfahrungen gebracht. Generell lässt sich feststellen, dass mit den in der Arbeitsgruppe gebräuchlichen Checklisten Informationen in erheblichem Umfang aus den Studien abgefragt werden, die Aufschluss über die methodische Studienqualität geben, ohne allerdings systematisch verwertet zu werden. Die Entscheidung, ob eine

Publikation mit ihren Ergebnissen in den HTA-Bericht einbezogen wird, wird anhand weniger, als relevant identifizierter Kriterien (QA-Items) gefällt. Die zeitaufwendige Beantwortung aller anderen Fragen wird lediglich dokumentiert, hat sonst aber keinen Einfluss auf Ergebnisse oder Schlussfolgerungen. Die zweite Erfahrung bezieht sich auf die Kontextdokumente: Bisherige Projektarbeiten haben gezeigt, dass es in aller Regel nicht möglich ist, im Ausland erarbeitete HTA-Berichte oder auch Praxisleitlinien 1:1 für die Entscheidungsfindung in Deutschland einzusetzen. Systemgegebene, ökonomische oder kulturelle Unterschiede machen eine kontextangepasste Neuinterpretation der Daten in jedem Fall erforderlich. Als hilfreich hat es sich jedoch erwiesen, eigene Literaturübersichten auf in ausländischen Kontextdokumenten enthaltenen Informationssynthesen aufzubauen – vorausgesetzt die methodische Qualität ist gegeben. Es wird somit nicht das komplette Kontextdokument weiter „verwertet“, sondern nur sein Kernstück, die systematische Literaturübersicht.

Diese Erfahrungen waren wegweisend für die im Folgenden vorgeschlagenen Vorgehensweisen. In Abwandlung der bestehenden Arbeitsanleitungen für HTA-Berichte in Deutschland sehen wir, auch angesichts der neueren methodischen Literatur, bei der Qualitätsbeurteilung der Primärmaterialien folgendes Kürzungspotential:

- Straffung und Kürzung des Bewertungsinstrumentariums – die Checkliste für Kontextdokumente wird verlassen, es wird eine neue, kürzere Checkliste für alle Sekundärpublikationen entwickelt.
- Die Checkliste 2a für alle Primärstudiendesigns (ausser für Diagnosestudien) wird ebenfalls verlassen und dafür werden kürzere, spezifische Instrumente für Prognosestudien und Diagnosestudien entwickelt. Zur Bewertung von Interventionsstudien schlagen wir die Verwendung eines modifizierten Jadad-Scores (Jadad et al. 1996) vor.
- Qualitätsbeurteilung durch einen Autor (anstelle zweier Autoren).

Ziel des „Critical Appraisal“ ist die Feststellung eines Mindestgrades von methodischer Qualität der Primärmaterialien, sodass eine relevante systematische Verzerrung der berichteten Studienergebnisse weitgehend ausgeschlossen werden kann. Angestrebt wird ein dichotomes Ergebnis der Beurteilung:

- Die Publikation wird aufgrund gravierender methodischer Mängel in der systematischen Literaturübersicht nicht berücksichtigt

oder

- die Ergebnisse der Studie / der Übersichtsarbeit werden in der systematischen Literaturübersicht berücksichtigt.

Hierzu wurden für die Verwendung in Kurz-HTAs Komponentenlisten für Literaturübersichten und Prognosestudien zusammengestellt und einer Reliabilitätstestung unterzogen. Eine Komponentenliste für Diagnosestudien wurde konzipiert. Da dieser Studientyp im Beispielthema nicht analysiert wurde, konnte das Instrument im Rah-

men des laufenden Projekts keiner Testung unterzogen werden und wird daher auch nicht berichtet.

Im Rahmen der protokollkonformen Themenbearbeitung sollen die Bewertungsinstrumente, um nach Möglichkeit Zeit einzusparen, nach dem Schema eines Stufenplanes eingesetzt werden. Zum Einschluss in die Übersicht müssen alle Fragen (bzw. die festgelegte Anzahl) in der Komponentenliste mit „Ja“ beantwortet werden – schon ein „Nein“ bedeutet automatisch den Ausschluss der Studie (wenn nicht sehr gute Gründe für eine Ausnahme vorliegen: z. B. wenn nicht publizierte fehlende Informationen durch Kommunikation mit den Autoren erfahren werden konnten).

Wird aufgrund einer beschränkten Verfügbarkeit von (qualitativ hochwertigen) Studienpublikationen mit modifizierten Ein- und Ausschlusskriterien gearbeitet, können die Checklisten auch als simple Score-Instrumente zur Dokumentation der Materialqualität verwendet werden - hierbei bleibt die "Ausschlussklausel" unberücksichtigt.

Im folgenden soll kurz die Zusammenstellung der Qualitätskriterien für die Stufenpläne „Literaturübersichten“, „Prognosestudien“ und „Diagnosestudien“ geschildert werden. Zur Entwicklung des „Jadad-Score“ verweisen wir auf die Originalpublikation von Jadad et al. (1996).

Stufenplan 1: Literaturübersichten

Konzeption:

Zusammengestellt werden soll ein Instrument, welches zu erkennen hilft, ob die Ergebnisse einer Literaturübersicht durch systematische Fehler verzerrt sind. Dabei macht es keinen Unterschied, ob es sich um eigenständig publizierte Übersichten handelt oder ob die Literaturübersicht Bestandteil eines Kontextdokumentes ist (HTA-Bericht, Leitlinie). Das Instrument soll einige, für die Validität von Reviewergebnissen essentiell relevante Komponenten abfragen. Dabei soll es im Sinne eines Stufenplans einsetzbar sein, d. h. in dem Moment, wo eine Frage verneinend beantwortet werden muss, kann die Bewertung abgebrochen werden. Das Ergebnis der Bewertung soll dichotom angegeben werden: Ein- oder Ausschluss der Publikation, eine Scorebildung und weitere Verwendung der Ergebnisse z. B. in Sensitivitätsanalysen oder als Gewichtungsfaktoren ist nicht vorgesehen.

Auswahl der Items:

Systematische Literaturübersichten sind ebenso wie Primärstudien als wissenschaftliche Arbeiten anzusehen. Eine Primärstudie zielt darauf ab, die Wirksamkeit eines Therapieverfahrens, die Auswirkungen eines Risikofaktors oder die Leistungsfähigkeit eines diagnostischen Testverfahrens ohne systematische Verzerrungen zu erfassen und zu berichten. Eine Literaturübersicht soll dagegen den wissenschaftlichen Kenntnisstand valide ohne systematische Verzerrungen berichten.

Für Literaturübersichten zu therapeutischen Verfahren gibt es inzwischen verbreitete und allgemein akzeptierte methodische Vorgehensweisen (Cochrane Collaboration; CRD 2001). Ein Verfahrensvorschlag zur Berichterstattung dieser Art Reviews liefert das „QUORUM-Statement“ (Moher 1999), welches in Analogie zum „CONSORT-Statement“ für RCTs zwei Adressatengruppen hat: die Wissenschaftler, die die Evidenz auswählen, bewerten und zusammenfassen sowie die Journalherausgeber bzw. Reviewer, die stellvertretend für die Leserschaft die Evidenz zur Kenntnis nehmen. Für den letztgenannten Adressatenkreis stellt das QUORUM-Statement eine Checkliste zur Verfügung, die es erleichtert, im Review die qualitätsrelevanten Passagen aufzufinden und gegebenenfalls ihr Fehlen festzustellen.

Die Methodik der Zusammenfassung von Studienergebnissen nicht-experimenteller, beobachtender Studiendesigns und diagnostischer Studien (Deeks 2001) ist nicht so bekannt und unterscheidet sich hauptsächlich in der Methodik der Meta-Analyse-Verfahren. Für Reviews / Meta-Analysen von Beobachtungsstudien wurde von der Arbeitsgruppe „Meta-Analysis of Observational Studies“ („MOOSE“) ebenfalls eine Checkliste zusammengestellt, die in Analogie zum QUORUM – Statement Qualitätsaspekte (35 Items in sechs methodischen Arbeitsschritten) von Reviews observatio-neller Studien nennt und auffinden hilft (Stroup et al. 2000).

Als „Itempool“ für den Stufenplan zur Qualitätsprüfung von Übersichtsarbeiten wurden herangezogen: das QUORUM-Statement, die Checkliste der „MOOSE“-Gruppe, die Übersichtsarbeit von Deeks zu diagnostischen Studien und die Checklisten 1a und 1b der deutschen HTA-Arbeitsgruppe (aus diesen in erster Linie die mit QA bezeichneten Items).

Unser Ziel war es, ein Instrument zu gewinnen, welches in der Lage ist, qualitativ hochwertige (interne und externe Validität) systematische Literaturübersichten von unsystematischen bzw. qualitativ unbefriedigenden zu trennen. Die Checkliste gliedert sich in fünf Untereinheiten. Die Aspekte, die in der ersten, mit „Filterfrage“ bezeichneten, Untereinheit A erfasst werden, dienen der Feststellung der externen Validität (Vergleichbarkeit der im Review eingeschlossenen Studienpopulationen, Interventionen und Zielgrößen mit denen der eigenen Fragestellung).

Ausgehend von der Überlegung, dass die interne Validität einer Literaturübersicht vor allem auf Vollständigkeit und hoher Qualität der eingeschlossenen Primärmaterialien beruht, sind die nächsten beiden Untereinheiten diesen beiden Aspekten gewidmet. In Einheit B „Informationsgewinnung“ wird in drei Fragen die Vollständigkeit der Übersicht, in Einheit C „Bewertung der Informationen“ die Feststellung der methodischen Qualität der Primärmaterialien dokumentiert. Die Untereinheit D ist zweigeteilt: Der erste Teil D1 bezieht sich auf alle Literaturübersichten und fragt nach der adäquaten Wertung und Diskussion der Primärmaterialien. Im zweiten Teil D2 wird nach drei Qualitätsaspekten von Meta-Analysen gefragt.

Für die protokollkonforme Themenbearbeitung ist vorgesehen, dass die Bewerten-den die Fragen des Instrumentes der Reihe nach beantworten. Sie sind so angeordnet, dass die jeweils weiter oben stehenden Fragen Voraussetzungen für weiter un-

ten abgefragte Aspekte bilden. Ist beispielsweise die Informationsrecherche unvollständig, erübrigt sich das Anlegen von strengen methodischen Ein- und Ausschlusskriterien, weil aufgrund der unvollständigen Ausgangsdaten systematische Verzerrungen der Ergebnisse nicht ausgeschlossen werden können.

Auf alle Fragen gibt es lediglich die Antwortmöglichkeiten „Ja“ und „Nein“ – der Reviewer ist somit gezwungen zu einer Entscheidung zu kommen. Für jede Untereinheit ist festgelegt, wieviel „Ja“-Antworten erforderlich sind, damit die Arbeit weiter bearbeitet und in den Review eingeschlossen wird. Wird die erforderliche Anzahl von „Ja“-Antworten in einer Kategorie nicht erreicht, kann die Bearbeitung der Studie abgeschlossen werden. Auf die Möglichkeit, eine Übersichtsarbeit trotz methodischer Mängel mit einer adäquaten Begründung einzuschliessen, sollte nur in Ausnahmefällen zurückgegriffen werden (z. B. wenn fehlende Angaben von den Studienautoren erhalten werden können oder wenn es sich bei der Arbeit um die einzige zu einer bestimmten Fragestellung handelt).

Wird die Themenbearbeitung mit modifizierten Ein- und Ausschlusskriterien vorgenommen, kann das Instrument als Score verwendet werden zur alleinigen Dokumentation der Studienqualität. Die Ausschlussvorgaben bleiben dann unbeachtet, es wird lediglich die Summe der mit "Ja" beantworteten Fragen erfasst.

Tabelle 9: Stufenplan 1

Stufenplan zur Validitätsprüfung von Literaturübersichten (incl. Meta-Analysen)		
Quelle:		
A Filterfrage: Ist das bearbeitete Thema relevant für die eigene Fragestellung?		
	ja	nein
Patientengruppe	<input type="checkbox"/>	<input type="checkbox"/>
Intervention	<input type="checkbox"/>	<input type="checkbox"/>
Zielgrößen	<input type="checkbox"/>	<input type="checkbox"/>
< 3 "Ja"-Antworten: Ausschluss	<input type="checkbox"/>	bzw. Begründung für Ausnahme:
B Informationsgewinnung vollständig und adäquat dokumentiert?		
Elektronische Informationsquellen (Datenbanken)	<input type="checkbox"/>	<input type="checkbox"/>
Suchstrategien	<input type="checkbox"/>	<input type="checkbox"/>
weitere Informationsquellen (Kongresse, Kontakte)	<input type="checkbox"/>	<input type="checkbox"/>
< 3 "Ja"-Antworten: Ausschluss	<input type="checkbox"/>	bzw. Begründung für Ausnahme:

C		
Bewertung der Informationen		
Transparente inhaltliche Ein- und Ausschlusskriterien?	<input type="checkbox"/>	<input type="checkbox"/>
Transparente methodische Ein- und Ausschlusskriterien?	<input type="checkbox"/>	<input type="checkbox"/>
Transparente Validitätsbeurteilung der Primärmaterialien?	<input type="checkbox"/>	<input type="checkbox"/>
< 3 "Ja"-Antworten: Ausschluss <input type="checkbox"/>		
bzw. Begründung für Ausnahme:		
D		
Diskussion/Schlussfolgerungen		
D1	Diskussion der methodischen Limitationen?	<input type="checkbox"/>
< 1 "Ja"-Antworten: Ausschluss <input type="checkbox"/>		
bzw. Begründung für Ausnahme:		
D2		
Informationssynthese / Schlussfolgerungen bei Meta-Analysen		
	Wahl des Meta-Analyse-Verfahrens begründet?	<input type="checkbox"/>
	Heterogenitätstest	<input type="checkbox"/>
	Sensitivitätsanalysen erwogen oder durchgeführt	<input type="checkbox"/>
< 3 "Ja"-Antworten: Ausschluss <input type="checkbox"/>		
bzw. Begründung für Ausnahme:		

Reliabilitätstestung:

Dem internationalen Standard entsprechend haben wir zunächst die Reliabilität der Checklisten in einer Pilot-Phase getestet, bevor wir die ermittelten Literaturartikel bewertet haben. Die Testung bezog sich auf die Verwendung des Instrumentes als "Stufenplan" mit der dichotomen Bewertung Einschluss oder Ausschluss.

Drei Reviews wurden von insgesamt sechs wissenschaftlichen Mitarbeitern des Instituts für Sozialmedizin der Universität Lübeck anhand der „Stufenpläne“ bewertet, eine Datenextraktion erfolgte nicht. Bei den sechs Mitarbeitern handelte es sich durchweg um Mediziner, zwei mit epidemiologischer Zusatzausbildung, zwei mit langjähriger theoretisch-methodischer Tätigkeit und zwei mit klinisch-wissenschaftlichem Hintergrund.

Zur Bewertung wurden aus dem Pool der zum Thema recherchierten Literatur folgende Arbeiten ausgewählt:

Homik J, Hailey D: Quantitative ultrasound for bone density measurement. September 1998, HTA 11

Prins et al.: The role of quantitative ultrasound in the assessment of bone: a review. Clin. Physiol. 18(1): 3-17; 1997

Glueer: Quantitativer Ultraschall. Status 1999. (Quantitative ultrasound. State of the art 1999) Radiologe; 39: 213-221; 1999

Die Arbeiten wurden gezielt ausgewählt, um ein möglichst breites Spektrum der vorhandenen methodischen Qualität abzudecken. Die Ergebnisse der Reliabilitätstestung sind gemeinsam mit den Ergebnissen der Testung des Instrumentes für Prognosestudien im Anschluss an dessen Vorstellung dargestellt.

Stufenplan 2: Prognosestudien:

Konzeption

Zu Beantwortung der zentralen Fragestellung für das Beispielthema war es im Rahmen des Kurz-HTA-Projektes erforderlich, ein Instrument für die Bewertung von Prognosestudien zusammenzustellen. Der Stufenplan sollte nach Möglichkeit Aspekte der internen (Design, Durchführung, Analyse) und externen (Kompatibilität von Population, Intervention, Outcomes) Validität berücksichtigen. Dabei soll es ebenfalls im Sinne eines Stufenplanes einsetzbar sein, d. h. in dem Moment, wo eine Frage verneinend beantwortet werden muss, kann die Bewertung abgebrochen werden. Das Ergebnis der Bewertung soll dichotom angegeben werden: Ein- oder Ausschluss der Publikation, eine Scorebildung und weitere Verwendung der Ergebnisse z. B. in Sensitivitätsanalysen oder als Gewichtungsfaktoren ist nicht vorgesehen.

Auswahl der Items

Klasse-I-Evidenz zur Beantwortung von Fragestellungen aus dem Bereich Prognose / Auswirkungen von Risikofaktoren liefern prospektive Kohortenstudien (CEBM 2001) bzw. Meta-Analysen mehrerer prospektiver Kohortenstudien. Während für RCTs eine Reihe von empirischen Untersuchungen Zusammenhänge zwischen Studiencharakteristika (wie Randomisierung, „Concealment of Allocation“ oder Verblindung) und Ergebnissen nachweisen konnten, liegen solche Informationen bisher für Prognosestudien nicht vor. Die Auswahl von Items zur Beurteilung interner Validität begründen sich daher auf theoretischen Überlegungen und den methodisch geschärften „gesunden Menschenverstand“ (Altman 2001).

Als „Itempool“ für den Stufenplan „Prognosestudien“ wurden die Arbeiten von Altman (2001), Laupacis et al. (1994) und die Checkliste 2a der deutschen HTA-Arbeitsgruppe (in erster Linie mit QA bezeichnete Items) herangezogen.

Das Instrument soll mit dem Ziel eingesetzt werden, solche prospektiven Kohortenstudien zu identifizieren, deren Ergebnisse, soweit beurteilbar, in keinem relevanten Ausmaß von systematischen Fehlern verzerrt sind.

Die Checkliste gliedert sich in fünf Untereinheiten. Weil die Items sowohl Aspekte der externen wie auch der internen Validität erfassen, wurde auf eine einführende Filterfrage verzichtet.

Die erste Untereinheit erfasst zwei Charakteristika der Studienpopulation: Die eindeutige Definition (und Anwendung) von Ein- und Ausschlusskriterien und die transparente Beschreibung der Studienpopulation unter Berücksichtigung möglicher Störfaktoren. Weil in den Kohortenstudien die Exposition nicht randomisiert zugeordnet werden kann, muss der Verteilung von möglichen Einflussfaktoren und Störgrößen (Confoundern) in den Untersuchungsgruppen größtmögliche Aufmerksamkeit gewidmet werden. Nur wenn die Verteilung dieser Faktoren bekannt und beschrieben ist, können sie (wie in Untereinheit E der Checkliste abgefragt) bei der Datenanalyse berücksichtigt werden. Untereinheit B ist der Erfassung der Exposition gewidmet. Hier werden nur methodische Aspekte angesprochen: Die Messung soll valide, reliabel und in allen Untersuchungsgruppen einer Studie gleichartig erfolgen. Das gleiche gilt für die in Untereinheit C abgefragte Erfassung der Zielgrößen. Hier wurde zusätzlich noch der inhaltliche Aspekt der Verwendung von patientennahen Outcomes aufgenommen. In Untereinheit D werden Dauer und Vollständigkeit der Nachbeobachtung erfasst – zwei Punkte, die, falls nicht erfüllt, häufig für die Ursache von systematisch verzerrten Studienergebnissen gehalten werden. In Untereinheit E wird, wie oben schon erwähnt, erfasst, ob bei der Datenanalyse wichtige Einflussgrößen berücksichtigt wurden.

Protokollkonforme Themenbearbeitung: Wie beim Stufenplan „Literaturübersicht“ ist auch bei dieser Checkliste vorgesehen, dass die Bewertenden die Fragen des Instrumentes der Reihe nach abarbeiten. Sie sind so angeordnet, dass die jeweils weiter oben angeordneten Fragen Voraussetzungen für weiter unten abgefragte Aspekte bilden. Ebenfalls wie beim Stufenplan „Literaturübersicht“ gibt es lediglich die Antwortmöglichkeiten „Ja“ und „Nein“ – der Reviewer ist somit gezwungen zu einer Entscheidung zu kommen. Für jede Untereinheit ist festgelegt, wieviel „Ja“- Antworten erforderlich sind, damit die Arbeit weiter bearbeitet und in den Review eingeschlossen wird. Wird die erforderliche Anzahl von „Ja“ in einer Kategorie nicht erreicht, kann die Bearbeitung der Studie abgeschlossen werden. Auf die Möglichkeit, eine Studie trotz methodischer Mängel mit einer adäquaten Begründung einzuschließen, sollte nur in Ausnahmefällen zurückgegriffen werden (z. B. wenn fehlende Angaben von den Studienautoren erhalten werden können oder wenn es sich bei der Arbeit um die einzige zu einer bestimmten Fragestellung handelt).

Themenbearbeitung nach modifizierten Ein- und Ausschlusskriterien: Auch dieses Instrument kann bei der modifizierten Bearbeitung als Score verwendet werden, zur alleinigen Dokumentation der Studienqualität. Die Ausschlussvorgaben bleiben dann unbeachtet, es wird lediglich die Summe der mit "Ja" beantworteten Fragen erfasst.

Das Instrument ist auf der folgenden Seite abgebildet.

Tabelle 10: Stufenplan 2

Stufenplan zur Validitätsprüfung von Prognosestudien		
Quelle:		
A Studienpopulation	ja	nein
Eindeutige Definition und Anwendung der Ein- und Ausschlusskriterien für die Studienteilnahme	<input type="checkbox"/>	<input type="checkbox"/>
Transparente Beschreibung der Studienpopulation unter Berücksichtigung möglicher Störfaktoren	<input type="checkbox"/>	<input type="checkbox"/>
< 2 "Ja"-Antworten: Ausschluss <input type="checkbox"/>		bzw. Begründung für Ausnahme:
B Exposition	ja	nein
Valide, reliable und gleichartige Erfassung der Exposition / Riskofaktoren	<input type="checkbox"/>	<input type="checkbox"/>
< 1 "Ja"-Antworten: Ausschluss <input type="checkbox"/>		bzw. Begründung für Ausnahme:
C Zielgröße	ja	nein
Valide, reliable und gleichartige Erfassung der Zielgrößen	<input type="checkbox"/>	<input type="checkbox"/>
Verwendung patientennaher Zielgrößen	<input type="checkbox"/>	<input type="checkbox"/>
< 2 "Ja"-Antworten: Ausschluss <input type="checkbox"/>		bzw. Begründung für Ausnahme:
D Follow-up	ja	nein
1. Ausreichend lange Nachbeobachtungsperiode	<input type="checkbox"/>	<input type="checkbox"/>
2. Vollständigkeit der Nachkontrolle (Follow-up-Rate i. d. R. $\geq 80\%$)	<input type="checkbox"/>	<input type="checkbox"/>
< 2 "Ja"-Antworten: Ausschluss <input type="checkbox"/>		bzw. Begründung für Ausnahme:
E Auswertung	ja	nein
1. Adjustierung für wichtige prognostische Faktoren	<input type="checkbox"/>	<input type="checkbox"/>
< 1 "Ja"-Antworten: Ausschluss <input type="checkbox"/>		bzw. Begründung für Ausnahme:

Reliabilitätstestung:

An der Reliabilitätstestung nahmen die oben bereits vorgestellten sechs Mitarbeiter des Institutes für Sozialmedizin teil. Die Testung bezog sich wieder auf die Verwendung des Instrumentes als "Stufenplan" mit der dichotomen Bewertung Einschluss oder Ausschluss.

Zur Bewertung wurden drei Kohortenstudien unterschiedlicher methodischer Qualität aus dem Pool der für das Beispielthema recherchierten Literatur ausgewählt:

- Gnudi et al.: Quantitative ultrasound and bone densitometry to evaluate the risk of nonspine fractures: a prospective study. *Osteoporos Int* 11: 518-523; 2000
- Huang et al.: Prediction of fracture risk by radiographic absorptiometry and quantitative ultrasound: a prospective study. *Calcif Tissue Int* 63 : 380-384; 1998
- Garnero et al.: Do markers of bone resorption add to bone mineral density and ultrasonographic heel measurement for the prediction of hip fracture in elderly women? The EPIDOS prospective study. *Osteoporos Int* 8 : 563-569; 1998

Die Teilnehmer wurden gebeten, alle sechs Publikationen mit Hilfe der Stufenpläne zu bewerten und sich für oder gegen den Einschluss der Studie in die Literaturübersicht zu entscheiden. Ausserdem wurden formlos Rückmeldungen und Kommentare zu Inhalten, Format und Praktikabilität der Stufenpläne erfasst.

In Tabelle 11 sind die Endergebnisse der Bewertung dargestellt.

Tabelle 11: Endergebnisse der Reliabilitätstestung, Stufenpläne 1 und 2.

Reliabilitätstestung Stufenpläne						
	Literaturübersichten			Prognosestudien		
	Homik, 1998	Prins, 1997	Glueer, 1999	Gnudi, 2000	Huang, 1998	Garnero, 1998
Tester 1	+	-	-	+	+	-
Tester 2	+	+	-	-	-	-
Tester 3	-	-	-	+	+	-
Tester 4	+	-	-	-	-	-
Tester 5	+	-	-	-	+	-
Tester 6	+	-	-	+	+	-

+ = Einschluss **-** = Ausschluss der Studie / Übersicht

Aus der Übersichtstabelle geht bereits hervor, dass die Bewertung der Literaturübersichten mithilfe des Stufenplanes zu weitgehend übereinstimmenden Ergebnissen kommt. Die Arbeit von Glueer wurde von allen Testern ausgeschlossen, die Arbeiten von Homik et Hailey und die Arbeit von Prins wurden lediglich einmal diskrepant beurteilt. Die Einzelauswertung (Tabelle 12) zeigte, dass die beiden abweichenden Voten von Kollegen abgegeben wurden, die erst über wenig Erfahrung in der Literatur-

bewertung verfügten. Die Sammlung von freien Kommentaren und Hinweisen ergab keinen Anhalt für einen unmittelbaren Modifikationsbedarf des Instrumentes.

Anders dagegen der Stufenplan für Prognosestudien. Hier wurde lediglich eine von drei Studien ohne Diskrepanzen beurteilt. (Garnero, Ausschluss der Studie). Bei der Studie von Gnudi stimmten drei Tester für, drei gegen den Einschluss, bei der Arbeit von Huang verteilten sich die Voten vier für, zwei gegen Einschluss. Die Einzelauswertung zeigte, dass die Tester in den Unterkategorien B (Expositionsmessung), C (Zielgrößen) und E (Auswertungsverfahren) kaum diskrepante Bewertungen abgaben. Nichtübereinstimmungen, die auch zur diskrepanten Gesamtbewertung führten, wurden den Unterkategorien A (Studienpopulation) und D (Follow-Up) dokumentiert. Die Rücksprache mit den Testern ergab folgendes Bild: Die übereinstimmende Ablehnung der Studie von Garnero et al. lag darin begründet, dass es sich hier um eine Fall-Kontroll-Studie und nicht um eine Kohortenstudie handelt. Die diskrepanten Voten bei den anderen Arbeiten wurden damit begründet, dass insbesondere die Frage nach der „ausreichenden“ Charakterisierung und der „ausreichend“ langen Beobachtungsdauer – also beides keine methodischen sondern inhaltliche Aspekte der Studien – für Personen (Mediziner) ohne thematische Einarbeitung bzw. ausführliches Begleitmaterial nicht zu beantworten sind. Die ablehnenden Voten wurden durchweg „aus Vorsicht“ abgegeben.

Für die weitere Arbeit mit den Stufenplänen bedeutet dies:

Stufenplan 1: Anfertigung von „Benutzerhinweisen“ für den Stufenplan; Einsatz des Instrumentes in weiteren Projekten unter vorläufiger Beibehaltung der Doppelbefundung, insbesondere wenn thematisch / methodisch noch unerfahrene Reviewer eingesetzt werden. Eine Validierung des Instrumentes sollte im Rahmen eines Projektes zur Validierung des gesamten Kurz-HTA unternommen werden.

Stufenplan 2: Anfertigung von ausführlichen Benutzerhinweisen, aus methodischer und inhaltlicher Perspektive. Wiederholung der Reliabilitätstestung. Falls das Ergebnis besser ausfällt, weiteres Prozedere wie mit Stufenplan 1. Weist die zweite Reliabilitätstestung erneut nicht-akzeptable Diskrepanzen auf, muss die Zusammenstellung eines neuen Instrumentes erwogen werden. Hierbei wäre eine klare Betonung von methodischen Qualitätsaspekten vorzunehmen, mehr inhaltliche Überlegungen müssten in den Selektionsprozess (vergl. SOP 2) verlagert werden. Wir weisen darauf hin, dass hier nicht die Bewertung der medizinischen Relevanz der Fragestellung zentral ist, sondern dass die formale Beschreibung der Planung, Durchführung, Auswertung der Studie und die klinischen Konsequenzen im Mittelpunkt stehen.

Tabelle 12 : Einzelauswertung der Reliabilitätstestung; Stufenplan 1 und Stufenplan 2

	Reliabilitätstestung Stufenplan 1: Literaturübersichten															Reliabilitätstestung Stufenplan 2: Prognosestudien																	
	1. Homik					2. Prins					3. Glueer					1. Gnudi					2. Huang					3. Garnero							
	G	a	b	c	d	G	a	b	c	d	G	a	b	c	d	G	A	B	C	D	E	G	A	B	C	D	E	G	A	B	C	D	E
Tester 1	+	+++	+++	+++	+	-	+++	+-	---	+	-	+-	---	---	-	+	++	+	++	++	+	+	++	+	++	++	+	-	--	+	++	++	+
Tester 2	+	+++	+++	+++	+	+	+++	+++	+++	+	-	-+	---	---+	-	-	+-	-	++	-+	+	-	--	+	++	+-	+	-	--	-	++	++	+
Tester 3	-	+++	+++	+++	+	-	+++	+-	+-	-	-	+-	---	---+	-	+	++	+	++	+-	+	+	++	+	++	-+	+	-	--	-	-+	-+	-
Tester 4	+	+++	+++	+++	+	-	---	+-	---	-	-	+++	---	---	-	-	+-	+	--	+-	+	-	--	-	-+	+-	-	-	+-	-	+-	--	+
Tester 5	+	+++	+++	+++	+	-	+++	--	+-	+	-	+-	---	---	-	-	+-	+	++	-+	+	+	++	+	++	++	+	-	--	+	++	++	+
Tester 6	+	+++	+++	+++	+	-	+++	+-	---	+	-	+-	---	---	-	+	++	+	++	++	+	+	++	+	++	++	+	-	--	+	++	-+	+

G Gesamtergebnis
 + Ja
 - Nein

Prognosestudien

- A Studienpopulation
- B Exposition
- C Zielgröße
- D Follow-up
- E Auswertung

Literaturübersichten

- a Filterfrage
- b Informationsgewinnung
- c Bewertung der Information
- d Diskussion/Schlussfolgerungen

Die zentrale Fragestellung für das Kurz-HTA „Quantitative Ultraschallverfahren“ fordert den Nachweis, dass das Ultraschallmessverfahren in der Lage ist, Personen mit einer (osteoporosebedingten) erhöhten Frakturgefährdung zu erkennen. Für die relevantesten / validesten Studien zur Beantwortung dieser zentralen („prognostischen“) Frage werden prospektive Kohortenstudien bzw. eine Meta-Analyse aus prospektiven Kohortenstudien gehalten (CEBM 2001).

Literaturrecherche und Vorselektion ergaben einen HTA-Bericht (Homik & Hailey, 1997), eine Literaturübersicht (Prins et al. 1997) und sieben Prognosestudien (prospektive Kohortenstudien). Die beiden erstgenannten (Übersichten) wurden mit dem Stufenplan 1, die Primärstudien mit dem Stufenplan 2 bewertet.

I C.3.3.3 Diskussion: Bewertung, eigene Erfahrungen

Weil die Reliabilitätstestung zumindest für den Stufenplan 2 auf Probleme bei der Übereinstimmung hinwies, wurde die Qualitätsbewertung der Literatur im Rahmen des Kurz-HTA in unverblindeter, unabhängiger Doppelbefundung vorgenommen. Bei dieser Vorgehensweise traten allerdings kaum Diskrepanzen auf, was wir uns durch die Tatsache erklären, dass beide Reviewer gründlich in das Themengebiet „Osteoporose“ eingearbeitet waren - Fehlinterpretationen der Fragen aus inhaltlicher Sicht daher nicht vorkamen. Drei Publikationen wurden wegen methodischer Mängel von der Berücksichtigung im Schnell-HTA ausgeschlossen (Tabelle 13).

Tabelle 13: Methodische Ausschlussgründe

Publikation	Ausschlussgrund / - gründe
Prins et al., 1997	∅ transparenten inhalt. und method. Ein- und Ausschlusskriterien, ∅ transparente Validitätsbeurteilung der Primärmaterialien
Garnero, 1998	∅ prospektive Studie, sondern eine eingenistete Fall-Kontroll-Analyse
Dretakis, 1998	∅ prospektive Studie, sondern eine Fall-Kontroll-Studie (Informationen zum prospektiven Teil erscheinen nur in der Diskussion)

Die unter dem Gliederungspunkt „Reliabilitätstestung“ skizzierten Konsequenzen sollten aber auf jeden Fall umgesetzt werden, auch wenn sich daraus für die derzeitige Themenbearbeitung keine Folgen mehr ergeben.

Abschließend muß noch einmal hervorgehoben werden, dass die Instrumente der Qualitätsbeurteilung – seien es systematische Auswertungsbögen oder detaillierte Checklisten mit Zusammenfassung der Qualität in einem Score – erheblich variieren. Angesichts dieser verschiedenen Möglichkeiten der Verfahrensweise sind transparente Arbeitsanleitungen der Qualitätsbewertung zumindest im Konsens zu harmonisieren, um die Vergleichbarkeit der Studienbewertungen zu gewährleisten. Die Standards zur Beurteilung von therapeutischen Verfahren haben, trotz der noch bestehenden Probleme, bereits ein hohes Niveau erreicht; die Standards zur Bewertung prognostischer und diagnostischer Studien, ökonomischer Analysen, Meta-Analysen

und Entscheidungsanalysen befinden sich dagegen in einem noch vergleichsweise frühen Entwicklungsstadium (Gibis 2001).

I C.3.3.4 Schlussfolgerung: SOP 3

Vorschlag für eine allgemeingültige Arbeitsanleitung:

Die hier vorgestellte SOP 3 „Literaturbewertung“ muss als Interimslösung interpretiert werden, da die Entwicklung der Stufenpläne noch nicht als abgeschlossen angesehen werden kann.

- | | |
|---|---------------|
| 1. Zur Bewertung von Literaturübersichten
(incl. HTA-Berichte oder Leitlinien) | Stufenplan 1 |
| 2. Zur Bewertung von Prognosestudien | Stufenplan 2* |
| 3. Zur Bewertung von Interventionsstudien | Jadad-Score |

*In Weiterentwicklung

I C.3.4 SOP 4: Datenextraktion

I C.3.4.1 Hintergrund: Status Quo (Ausführliches HTA)

Die Informationssynthese bildet das Kernstück jeder systematischen Literaturübersicht. In ihr werden Studiendetails und –ergebnisse präsentiert, die zwar in der Diskussion noch gewertet oder in einen bestimmten Kontext gesetzt werden, aber schließlich die Basis für die Schlussfolgerungen bilden.

Zur Vorbereitung der Informationssynthese werden alle für die Fragestellung der HTA relevanten Daten aus den Primärstudien „extrahiert“. Mit „extrahieren“ ist hier gemeint, dass aus jeder Studie die gleichen relevanten Informationen und Ergebnisse herausgelesen und übersichtlich tabellarisch präsentiert werden. Das Herauslesen von Informationen aus Studien ist prinzipiell ein sehr subjektiver Prozess und damit anfällig für systematische Verzerrungen. Ziel dieses SOP 4 ist es, den Vorgang der Datenextraktion weitgehend zu objektivieren und transparent zu dokumentieren. Es umfasst die Komponenten:

- Inhalte der Datenextraktion
- Konstruktion des Datenextraktionsformulars
- Methodik der Datenextraktion
- Umgang mit Problemen bei der Datenextraktion

Eine Datenextraktion wird im Rahmen eines systematischen Review in jedem Fall vorgenommen, unabhängig davon, ob eine quantitative (Meta-Analyse) oder eine qualitative (narrative) Informationssynthese vorgenommen werden soll.

Das Toolkit für ausführliche HTAs gibt keine spezielle Anleitung zur Konstruktion von Formularen zur Datenextraktion. Die einzigen beiden Vorgaben sind die Forderung, die Datenextraktion durch zwei unabhängige Personen durchführen zu lassen sowie die Aufforderung, ein Datenextraktionsformular mit definierten Fragestellungen zu verwenden.

Eine im großen und ganzen allgemein akzeptierte Vorgehensweise bei der Datenextraktion im Rahmen von systematischen Reviews sind dem Handbuch der Cochrane Collaboration (Cochrane Reviewers' Handbook 4.1.3, updated October 2001, www.cochrane.dk/cochrane/handbook/handbook.htm), der „Guidance for Systematic Reviews“ des CRD (Undertaking Systematic Reviews of Research on Effectiveness, CRD Report Number 4, 2nd Edition, March 2001, www.york.ac.uk/inst/crd/report4.html) oder dem Buch Systematic Reviews (Chalmers I, Altman DG, eds. Systematic Reviews. London: BMJ 1995) zu entnehmen. Zu diesen „quasi-Standards“ gehören inhaltliche, design-relevante und methodisch-prozedurale Aspekte sowie Vorschläge zum Umgang mit häufig auftretenden Problemen.

Inhalte

Es kann aufgrund der Heterogenität von HTA-relevanten Fragestellungen und Forschungsdesigns keinen allgemeingültigen „Katalog“ für Daten geben, die aus den Primärpublikationen zu extrahieren sind. Wie schon bei den Checklisten für die kritische Studienbewertung kann auch bei der Datenextraktion ein „Zuviel“ an wenig relevanten Informationen nicht nur einen erheblichen Zeitaufwand bedeuten, sondern auch die eigentlich gewünschten Kernaussagen eher verschleiern. Ein „Zuwenig“ an Informationen birgt dagegen die Gefahr, wichtige, vor allem auch für die Beurteilung der externen Validität der Reviewergebnisse erforderliche Informationen nicht zu berücksichtigen. Es ist zu bedenken, dass Informationen zu den Primärstudien, die nicht als Extrakte im Review angeboten werden, für den Leser / Nutzer der Übersicht nur mit erheblichem Aufwand zugänglich sind.

Die Festlegung der aus den Primärpublikationen zu extrahierenden Informationen muss sich daher sehr eng an den vordefinierten (vergl. SOP 1) Fragestellungen orientieren. Die geplante Ergebnisanalyse gibt somit die Auswahl der zu extrahierenden Informationen vor. Insbesondere sind auch solche Informationen einzuschließen, die für eine Bewertung der klinischen und der statistischen Homogenität der Studien erforderlich sind bzw. die für mögliche Sensitivitätsanalysen von Bedeutung sein können (vergl. SOP 5). Einige Schlüsselkomponenten sind jedoch für (fast) alle Datenextraktionen vorzusehen (modifiziert nach CRD März 2001).

Tabelle 14: Schlüsselkomponenten für die Datenextraktion (modifiziert nach CRD 2001)

Datenkategorie	Inhalte
Allgemeine Informationen	<ul style="list-style-type: none"> • Bearbeitungsdatum, Revieweridentifikation • Autoren, Titel, bibliographische Details oder andere Angaben zur Identifikation der Studie • freies Feld für Bemerkungen oder Kommentare
Spezifische Informationen - Studiencharakteristika	<ul style="list-style-type: none"> • Studientyp (- und damit Verifikation des Einschlusskriteriums), methodische Studienqualität • Charakteristika der Studienpopulation und des Versorgungssettings • für Wirksamkeitsstudien: Interventions- und Kontrollbedingungen • für diagnostische Studien: Test- und Goldstandardverfahren • für Prognosestudien: Expositionsdefinition, -abstufungen • Zielgrößen (Outcomes) • freies Feld für Bemerkungen oder Kommentare
Spezifische Informationen – Ergebnisse	<ul style="list-style-type: none"> • Nachbeobachtungsdauer, Studienabbrucher, fehlende Daten • Ergebnisse (z.B. diskrete und kontinuierliche Daten, Überlebenszeitkurven, Sensitivität, Spezifität u. a.) • ggf. Effektstärken • freies Feld für Bemerkungen oder Kommentare
Hinweise zur Formularbenutzung / Kodierhinweise	

Diese Schlüsselkomponenten sollen nur grob die Rahmenbedingungen für eine Datenextraktion skizzieren, je nach Fragestellung müssen die einzelnen Punkte entsprechend differenziert und spezifiziert werden.

Design der Datenextraktionsformulare

Prinzipiell bieten sich für die Konstruktion von Datenextraktionsformularen zwei Vorgehensweisen an:

Formulare können in Papierform oder als einfache Textverarbeitungsdokumente z. B. tabellarisch die extrahierten Informationen aufnehmen. Sollen Berechnungen mit den Daten vorgenommen werden, ist eine erneute Eingabe der Überführung in eine Datenbank oder ein Statistikprogramm erforderlich.

Mithilfe moderner Datenbankprogramme z. B. MS ACCESS[®] lassen sich elektronische Formulare konstruieren, deren Daten dann auch elektronisch weiterverarbeitbar sind und sich in unterschiedlichen Formaten präsentieren lassen. Der Einsatz solcher Programme sollte insbesondere dann erwogen werden, wenn es sich um sehr viele Studien handelt, deren Einzeldaten computerisiert sehr viel einfacher zu überblicken sind oder wenn die Durchführung einer Meta-Analyse geplant ist. Das Programm RevMan[®] der Cochrane Collaboration (Freeware) ist ein spezielles Hilfsprogramm zur Erstellung eines systematischen Review nach den methodischen Vorgaben der Cochrane Collaboration.

Wegen der notwendigen Anpassung an die genaue Berichtsfragestellung muss das Datenextraktionsformular für jeden Bericht neu erstellt werden. Ist eine Arbeitsgruppe allerdings häufiger mit der Erstellung von Reviews befasst, können bereits vorhandene Formulare auf weitere Berichte übertragen werden, wenn die Strukturen der Studien sich ähneln. In dieser Situation bietet es sich an, eine einheitliche Struktur für Datenextraktionsformulare und Kodierregeln zu entwickeln, die beibehalten werden und bei den Reviewern Fehler vermeiden helfen.

Einige internationale HTA-Institutionen fordern auch ein „pilot testing“ der Formulare, um eine einheitliche, reproduzierbare und vollständige Datenextraktion zu gewährleisten.

Methoden der Datenextraktion:

Richtigkeit, Genauigkeit und - wenn mehrere Reviewer in den Extraktionsprozess eingebunden werden - Übereinstimmung sind extrem wichtig bei der Datenextraktion. Daher schlägt das CRD vor, folgende Mindestforderungen einzuhalten, um zufällige und systematische Fehler zu vermeiden:

- Klare Instruktionen und Richtlinien; Anleitung zur Vorgehensweise direkt auf das Formular schreiben
- Mindestens zwei Datenerheber
- Training und praktische Übungen für Reviewer
- Unstimmigkeiten bei der Datenextraktion und Datenmanipulation müssen diskutiert werden, ansonsten muß eine Sensitivitätsanalyse durchgeführt werden
- Die relevanten Studien und Bewertungen sollten vor der Datensynthese in einem „Meeting“ der HTA-Arbeitsgruppe diskutiert werden

Häufige Probleme bei der Datenextraktion

Vom CRD werden zusätzlich zu den Verfahrensvorschlägen auch einige Vorschläge zum Umgang mit Problemen bei der Datenextraktion gemacht.

Ein relativ häufig angetroffenes Problem sind die Mehrfachpublikationen, d. h. Arbeitsgruppen publizieren die gleichen Forschungsergebnisse mehrfach in unterschiedlichen Publikationsorganen, möglicherweise mit geringfügig differierender Schwerpunktsetzung. Davon abzugrenzen sind sogenannte serielle Publikationen, die die Ergebnisse einer Studie in einer Serie von Publikationen mit unterschiedlich langen Nachbeobachtungszeiträumen berichten. Selbstverständlich sollten die Ergebnisse einer Studie nur einmal in eine systematische Übersicht einfließen, insbesondere in die Berechnung von Meta-Analysen. Aus einer Reihe von seriellen Publi-

kationen ist nur die neueste Publikation mit den bis dahin endgültigen Ergebnissen zu verwenden.

Können einer Publikation nicht alle erforderlichen Informationen entnommen werden, sollte immer versucht werden, über Kontaktaufnahme zu den Autoren die erforderlichen Informationen zu erhalten.

In Ausnahmefällen können auch statistische Verfahren herangezogen werden, um fehlende Parameter zu berechnen (z. B. Gruppengröße aus berichteten Behandlungserfolgen oder Hazard Ratios aus Überlebenszeitkurven). Diese Verfahren kommen aber nur in Vorbereitung von Meta-Analysen zur Anwendung. Ihre Auswirkungen auf das Gesamtergebnis sollten dann auf jeden Fall in Sensitivitätsanalysen überprüft werden.

I C.3.4.2 Methoden: Modifikations- / Kürzungspotential

Ein sorgfältiges Vorgehen bei der Datenextraktion ist unerlässlich, da sonst verfälschte Resultate zu erwarten sind. Auch die Konstruktion der Formulare sollte nach den o. g. Kriterien erfolgen und bietet, verglichen mit internationalen Standards, kein Kürzungspotential.

Denkbar wäre die Reduzierung des Zeitbedarfs für diesen Arbeitsschritt nur, indem die Literaturanalyse sehr eng auf die Reviewfragestellung fokussiert wird und indem die Datenextraktion nur durch eine Person durchgeführt wird. Die enge Fokussierung setzt allerdings eine sorgfältige Abstimmung mit den Nutzern und Fachgebietsexperten voraus (vergl. SOP 1), da Nacharbeiten – für die die komplette Literatur erneut durchgearbeitet werden müsste - einen erheblichen Zeitaufwand bedeuten würden.

Der Verzicht auf eine zweite Person spart personelle Ressourcen und durch die entfallende Klärung von Diskrepanzen auch Zeit. Allerdings muß dann diese Datenerhebung durch einen sehr erfahrenen Reviewer erfolgen bzw. erfordert eine sorgfältige Einarbeitung.

I C.3.4.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt

Wir fokussierten unser Formular auf die zentrale Fragestellung unseres Berichtes, d. h. „Wie valide läßt sich das Frakturrisiko mit Hilfe der quantitativen Ultraschallverfahren im Vergleich zum DXA-Verfahren abschätzen?“

Das Datenextraktionsformular wurde nach den oben angeführten inhaltlichen Kriterien zusammengestellt. Die Angaben wurden tabellarisch auf Papier bzw. im Textverarbeitungsprogramm erfasst. Auf die Aufnahme der Informationen in eine elektronische Datenbank wurde verzichtet, da es sich nur um sieben (bzw. fünf) zu bearbeitende Studien handelte und weil keine meta-analytische Datenzusammenfassung geplant war.

Bei dem einzuschließenden Studientyp handelte es sich um prospektive Studien, die in einer Patienten- / Probandenpopulation zu Studienbeginn eine Feststellung der Exposition (Bestimmung von „Knochendichte“parametern per Ultraschallverfahren) vornehmen und nach Ablauf einer Nachbeobachtungsdauer von mehr als 24 Monaten Frakturhäufigkeit in Abhängigkeit von der Exposition bestimmen.

Ausgeschlossen wurden Studien an Patienten mit sekundären Osteoporosen, in vitro Studien, Tierstudien, Fallberichte sowie Arbeiten, deren Originalsprache nicht Englisch, Deutsch, Spanisch oder Französisch war.

Folgende Informationen wurden extrahiert:

Tabelle 15: Extrahierte Daten für das Kurz-HTA "Ultraschall"

Allgemeine Informationen	<ul style="list-style-type: none"> • Erstautor • Erscheinungsjahr (diese beiden Angaben ermöglichen Zuordnung der Studie zur Publikation in der Referenzliste) • Studiendesign (damit Überprüfung des Einschlusskriteriums) • Besonderheiten / Kommentare
Spezifische Informationen - Charakteristika der Studienpopulation	<ul style="list-style-type: none"> • Größe (Anzahl Studienteilnehmer) • Alter • Geschlecht, Ethnizität • Beruf • Ko(morbidität)
	<ul style="list-style-type: none"> • Einschlusskriterien / Ausschlusskriterien • Land der Studiendurchführung
Expositionsmessung	<ul style="list-style-type: none"> • Gerätetypen und -namen • Messorte / -verfahren • Präzisionswerte in CV % • Ergebnisse: BUA, SOS, VOS, UTV
Zielgrößen	<ul style="list-style-type: none"> • Frakturen (Anzahl, Häufigkeit, Lokalisation etc.)
Spezifische Informationen – Ergebnisse	<ul style="list-style-type: none"> • relative Frakturrisiken (RR) bei exponierten Personen (niedrige QUS-Parameter) in Vergleich zu nicht-exponierten Personen (unauffällige QUS Parameter) • Adjustierungen für Confounder, Effektmodifikatoren

Das Formblatt sowie die Ergebnisse der Datenextraktion, die von zwei Personen unabhängig voneinander durchgeführt wurde, können bei den Autoren angefordert werden bzw. finden sich im Text.

I C.3.4.4 Diskussion: Bewertung, eigene Erfahrungen

Insgesamt wurden die oben genannten Angaben aus fünf Primärstudien (Huang 1998; Dargent-Molina 1999; Gnudi 2000; Stewart 1999; Pluijm 1999) und einem HTA-Bericht mit integrierter systematischer Literaturübersicht extrahiert und dokumentiert. Trotz der überschaubaren Menge an Literatur hat der gesamte Prozess der Datenextraktion 7,5 Tage Arbeitszeit in Anspruch genommen, hierbei fielen auf die wissenschaftliche Mitarbeiterin 5 Tage und auf den „Doppelbewerter“ noch einmal 2,5 Tage. Auf eine Doppelbefundung sollte aus Gründen der Personaleinarbeitung im Rahmen dieses Projektes nicht verzichtet werden.

Bei der überschaubaren Menge an Literatur und der durch die definierte Fragestellung sehr begrenzten Menge an zu verwaltenden Informationen war es sicherlich adäquat, auf ihre Erfassung in ein Datenbankprogramm zu verzichten. Für die Zukunft sollte diese Methode jedoch unbedingt in Erwägung gezogen werden. Mit einem zeitlichen Mehraufwand ist nicht zu rechnen, da gebräuchliche Programme über ausreichend Funktionalität für diese Aufgabe verfügen, sodass keine Programmierarbeiten im Rahmen der Formularerstellung anfallen werden.

Weitere Erfahrungen lassen sich an dieser Stelle nicht berichten, da sich das hier beschriebene Prozedere nicht gegen einen für das ausführliche HTA festgeschriebenen Standard kontrastieren lässt – es ist keiner vorhanden!

I C.3.4.5 Schlussfolgerung: SOP 4

Vorschlag für eine allgemeingültige Arbeitsanleitung

1. Identifikation der relevanten Informationen aus den Studien:

- Welche Angaben werden gebraucht, um die Reviewfragestellungen zu beantworten? Welche Aspekte könnten Studienheterogenität bedingen bzw. für Sensitivitätsanalysen interessant sein?

2. Definition (und ggf. Abstimmung) der Inhalte für das Extraktionsformular, dabei unverzichtbar:

- Allgemeine Informationen
- (Erst)Autor
- Erscheinungsjahr der Studie
- Studiendesign
- Spezifische Informationen
- Studiencharakteristika
 - Studienpopulation
 - Intervention (Exposition oder Testcharakteristika)
 - Zielgrößen
- Messung der Zielgröße und Ergebnisse
 - Zeitraum des follow-up
 - Drop-Outs

Fehlende Daten
Ergebnisse (diskrete Daten: Ereignisse, Anzahl; kontinuierliche Daten:
Mittelwert, Standardfehler, Standardabweichung; P-Wert);
Überlebensdaten; Zusammenhangsmaße)

- Raum für Kommentare / Anmerkungen

- 3. Konstruktion des Datenextraktionsformulars** „auf Papier“ oder elektronisch (z. B. RevMan© oder Access© , dabei Integration der Instruktionen/Anweisungen für den/die Datenerheber.

I C.3.5 SOP 5: Struktur und Darstellungsmöglichkeiten der Informationssynthese

I C.3.5.1 Hintergrund: Status Quo (ausführliches HTA)

Für die Informationssynthese, d. h. die zusammenfassende Darstellung der Ergebnisse eines Reviews können prinzipiell zwei unterschiedliche Herangehensweisen gewählt werden. Eine qualitative Informationssynthese präsentiert die Ergebnisse der eingeschlossenen Einzelstudien in einer übersichtlichen, z. B. tabellarischen Form. In der quantitativen Informationszusammenfassung wird mit Hilfe statistischer Verfahren (Meta-Analysen) ein „gepooltes“ Ergebnis aus den Ergebnissen mehrerer Einzelstudien errechnet. Im Prinzip unternimmt jeder systematische Review eine qualitative Informationssynthese, die unter bestimmten Voraussetzungen um die Berechnung eines gepoolten Ergebnisses ergänzt werden kann.

Bevor die Voraussetzungen für die Verfahrenswahl weiter vertieft werden, ist es sinnvoll, einige wenige Schlüsselbegriffe, die im Zusammenhang mit Informationssynthesen häufig verwendet werden, kurz zu definieren.

In den bisher erarbeiteten HTA-Projekten der deutschen Arbeitsgruppe wurden in der Mehrzahl qualitative Verfahren der Informationssynthese gewählt. Nur in Ausnahmefällen kamen quantitative Verfahren zur Anwendung. Das "Toolkit" gibt wenig konkrete Hinweise, welches Verfahren verwendet werden sollte. Somit kann auch zu diesem Punkt keine explizite Kontrastierung zum Standardverfahren erfolgen.

Tabelle 16: Glossar der Informationssynthese (nach CRD 2001)

Qualitative (deskriptive) Informationssynthese	<p>Nicht-quantitatives Verfahren zur Darstellung der gesammelten Evidenz, zur Abschätzung ihrer Wertigkeit und zur Vorbereitung von quantitativen Syntheseverfahren.</p> <p>Die qualitative Darstellung erlaubt eine qualitative Bewertung von Studiencharakteristika, -qualität und -ergebnissen (Heterogenität). In Situationen mit nur wenigen Studien, die große und konsistente Effekte nachweisen konnten, ist die qualitative Informationszusammenfassung ausreichend als Basis für Schlussfolgerungen.</p>
Quantitative Informationssynthese	<p>Eine Form der Informationszusammenfassung, bei der eine Reihe von statistischen Verfahren eingesetzt werden, um die Ergebnisse von Einzelstudien zu einem Gesamtergebnis zu kombinieren (Meta-Analyse), das Ausmaß der Heterogenität von Studien zu bewerten und Aspekte wie den Einfluss von „Publication Bias“ quantitativ zu bewerten. Mithilfe der Meta-Analyse wird ein kombinierter Effektschätzer und sein Konfidenzintervall berechnet.</p>
Heterogenität	<p>Die Variabilität bzw. Unterschiede zwischen den Studien, welche Schlüsselcharakteristika (klinische H.), Studienqualität (methodische H.) oder Effekte (H. der Ergebnisse) betreffen kann. Statistische Testverfahren werden eingesetzt, um festzustellen, ob die Heterogenität von Studienergebnissen größer ist als aufgrund von Zufallsschwankungen zu erwarten wäre.</p>
Homogenität	<p>Als Homogenität wird der Grad der „Ähnlichkeit“ von Studien bezeichnet. Studien werden als statistisch „homogen“ bezeichnet, wenn ihre Ergebnisse nicht stärker variieren, als es aus Zufallsgründen zu erwarten wäre.</p>
Sensitivitäts- Analysen	<p>Sensitivitätsanalysen untersuchen, inwieweit die Ergebnisse von systematischen Reviews in Abhängigkeit von unsicheren Entscheidungen oder Annahmen variieren.</p>
„Publication bias“	<p>Eine systematische Beeinflussung der Literaturrecherche (-auswahl), die dadurch entsteht, dass die Wahrscheinlichkeit der Publikation von Studienergebnissen von der Signifikanz ihrer Ergebnisse abhängig ist. (Eine Studie, deren Ergebnisse keinen signifikanten Effekt einer Intervention zeigen konnte, ist meistens schwieriger zu publizieren). Systematische Reviews, in denen keine Anstrengungen unternommen wurden, um unpublizierte Studien aufzufinden, überschätzen möglicherweise die wahre Wirksamkeit von Interventionen. Auf manchen Themengebieten (z. B. Alternativmedizinische Verfahren) wird diese Art von systematischer Verzerrung eher häufig angetroffen.</p>

Qualitative Verfahren:

Folgende aufeinander aufbauende Arbeitsschritte wurden in der Komponentenanalyse als relevant für die Informationssynthese identifiziert:

- - narrative Zusammenfassung (unterste Stufe)
- - Ergebnistabellen
- - qualitative Ergebniszusammenfassungen mit den unverzichtbaren Komponenten:
 - a) Studienpopulation(en)
 - b) Interventionen
 - c) Setting und Modifikationsfaktoren
 - d) Outcomes (Art, Robustheit, Bedeutung)
 - e) methodische Limitationen

Im Gegensatz zu den rein wissenschaftlichen systematischen Literaturübersichten und Metaanalysen (z. B. denjenigen der Cochrane Collaboration) muss das HTA-Verfahren laut Definition einen Zweck erfüllen: es bildet die Grundlage für eine gesundheitspolitische, strategische Entscheidung. D. h. im Rahmen eines HTA-Verfahrens wird es immer darum gehen, Studienergebnisse und Literatur von sehr heterogener Qualität zu berücksichtigen und in die Entscheidungsgrundlage zu integrieren. Schlussfolgerungen eines HTA-Berichtes können somit, im Gegensatz zur "ungezielten" Literaturübersicht nicht den Studienergebnissen folgen, sondern werden durch den problematischen Kontext (die Policy Question) und die Arbeitsfragen vorgegeben.

Qualitative Informationszusammenfassungen sollen die verfügbaren Informationen / Daten in einer geeigneten Form präsentieren, die es dem Leser erspart, die Originalliteratur selber durchzuarbeiten. Hierzu werden alle (für die HTA-Fragestellungen) relevanten Details der einzelnen einzubeziehenden Originalien (in diesem Zusammenhang gilt auch ein bereits existierender ausländischer HTA-Bericht oder ein Review als Originalarbeit) dargestellt aber noch nicht kommentiert oder diskutiert.

- Narrative Darstellung als Text: Gebräuchliche Darstellungsmethode, allerdings im Kontext von HTA-Verfahren mit einer Reihe von Nachteilen behaftet:

Unübersichtlichkeit: Sollen mehrere Aspekte pro Studie aus der Literatur herausgearbeitet werden, wird eine reine Textdarstellung mühsam lesbar (zum Vergleich der Einzelaspekte muss über weite Passagen im Text gesprungen werden) oder führt zu Redundanzen (da dieselbe Publikation mehrfach angesprochen werden muss). Kernaussagen sind schlecht erkennbar.

Subjektivität: Schon durch Gestaltung der Reihenfolge, Wortwahl oder häufige Bezugnahmen lässt sich eine textliche Zusammenfassung kaum wertneutral gestalten.

Eine rein narrative Informationssynthese ist unseres Erachtens nur dann vertretbar, wenn sich eine sehr eng umschriebene Fragestellung (methodisch hochwertige Studien / Primärmaterialien, hochspezifische Patientengruppe, gut definiertes - abgrenzbares - Verfahren, maximal zwei interessierende Outcomes) auf dem Boden von wenigen, heterogenen Primärmaterialien (z. B. ein systematischer Review oder HTA-Bericht + eine Einzelstudie) beantwortet werden müssen. Aber selbst dann ist zu fragen, ob nicht eine tabellarische Darstellung für die Verständlichkeit von Vorteil wäre.

- Ergebnistabellen: Die Darstellung von Daten in Ergebnistabellen kann die oben als Nachteile angeführten Aspekte als Vorteile verbuchen.

Tabellarische Darstellungen sind (in der Regel) übersichtlich und ermöglichen das Erkennen relevanter Informationen auf einen Blick. Die Gefahr der unbeabsichtigten Wertung ist bei der tabellarischen Darstellung nicht so groß und auch aus der zeitlichen Perspektive bietet sich die tabellarische Informationssynthese an: Die Datenextraktionsformulare (s. SOP 4) können direkt eingesetzt oder weiterverarbeitet werden. Allein dargeboten können auch Ergebnistabellen zu Fehlinterpretationen führen - im Text ist zumindest zu erläutern, wie die Tabelleninhalte erarbeitet wurden.

Für den Regelfall eines HTA-Schnellverfahrens halten wir eine kombinierte tabellarische und erläuternd sprachliche Darstellung der Informationssynthese für sinnvoll. Folgende Aspekte sind dabei regelhaft anzusprechen (Die Inhalte werden als regelhaft bezeichnet, obwohl strenggenommen die Arbeitsfragestellungen die relevanten Informationskomponenten determinieren. Die hier angeführten Aspekte sind jedoch im Rahmen von HTA-Gutachten fast immer von Interesse.):

a) Studienpopulation(en), b) Interventionen, c) Setting und Modifikationsfaktoren. Das Darstellen und Erläutern dieser Aspekte in einer qualitativen Informationssynthese bereitet die inhaltliche Diskussion der Ergebnisse des HTA-Berichtes vor. Sind die Ergebnisse für die eigene Zielpopulation, für die zur Debatte stehende Intervention im gegebenen Versorgungskontext anwendbar? Diese Fragen wird die Diskussion beantworten müssen - die erforderlichen Informationen werden in der Synthese bereitgestellt.

b) Outcomes (Art, Robustheit): unter diesem Aspekt werden die eigentlichen Ergebnisse der Einzelstudien aufgeführt. Die Art der angegebenen Outcomes wird ebenso beschrieben wie die Größe der beobachteten Effekte (z. B. Unterschiede zwischen Untersuchungsgruppen) und ihre statistische und klinische Relevanz. Zusammenfassungen werden höchstens deskriptiv vorgenommen (z. B. "die erzielte Blutdrucksenkung lag in vier der sechs Studien zwischen 10 und 15 mm Hg, in zwei Studien wurde keine Senkung berichtet").

c) Methodische Qualität der Primärmaterialien: Dieser Aspekt soll die Diskussion der Evidenz, ein mögliches Grading und die Schlussfolgerungen vorbereiten. Die metho-

dischen Aspekte der im Rahmen von SOP 2 definierten Einschlusskriterien für Primärstudien werden hier aufgegriffen und wertneutral berichtet. Falls die Qualitätsbewertung der Primärmaterialien nicht nur für eine Ein- / Ausschlussentscheidung intendiert war, sondern eine differenzierte Beschreibung ermöglicht, werden ihre Ergebnisse hier referiert (ggf. mit Punktescore - vergl. SOP 3).

In der Mehrzahl der Fragestellungen für Kurz-HTAs ist eine derartige Zusammenfassung der Studiencharakteristika und Ergebnisse durchaus ausreichend aussagekräftig, d. h. erlaubt die Formulierung von evidenzbasierten Schlussfolgerungen. Meistens handelt es sich hierbei um Technologien, die sich im Übergang von experimenteller Phase zur frühen Implementationsphase befinden oder bislang methodisch nur mit schwachen Studiendesigns evaluiert worden sind.

Ausnahmsweise kann es jedoch wünschenswert oder erforderlich sein, auch im Rahmen von Kurz-HTA-Verfahren eine meta-analytische Zusammenfassung von Studienergebnissen vorzunehmen.

Quantitative Informationssynthesen

Im Rahmen von HTA-Berichten werden quantitative Informationssynthesen immer nur für eng umgrenzte Aspekte der Fragestellung eingesetzt werden können. Charakteristischerweise handelt es sich dabei eigentlich immer um Efficacy-Aspekte, von therapeutischen, aber auch von diagnostischen Verfahren. Dies hat einfache methodische Gründe: Zur Durchführung von Meta-Analysen sind mehrere klinisch und methodisch homogene (s. u.) Studien zur Fragestellung erforderlich - diese finden sich häufiger als Untersuchungen unter Alltagsbedingungen. Meta-Analysen sind vor allem dann hilfreich, wenn zur Fragestellung eine Reihe von Studien publiziert wurden, die jedoch, jede für sich gesehen, zu wenig Teilnehmer enthalten, um zu statistischen aussagekräftigen Ergebnissen zu kommen. Auch wenn mehrere Studien vorliegen, deren Ergebnisse inkonsistent sind, d. h. es gibt sowohl positive als auch negative Studienergebnisse, können Meta-Analysen manchmal Aufschluss über die tatsächliche Effektivität geben.

Ausgedehnte Ausführungen zu quantitativen Syntheseverfahren würden den Rahmen dieses Methodenmanuals sprengen, es sollen jedoch prinzipielle Arbeitsschritte zur Erstellung einer Meta-Analyse kurz angesprochen und einige Hinweise für ihren Einsatz im Rahmen eines Schnellverfahrens gegeben werden (nach Burls et al. 2000).

a) Entscheidung: Sollte eine Meta-Analyse durchgeführt werden? Diese Entscheidung erfordert eine Beurteilung, ob die eingeschlossenen Studien sich ähnlich genug sind, um ihre Ergebnisse zu einem Gesamtergebnis zusammenzuführen. Diese "Ähnlichkeit" muss für die folgenden Dimensionen dokumentiert werden:

- Studienqualität (Studien niedriger methodischer Qualität tendieren dazu, deutlichere Effekte zu produzieren).

- Klinische Homo- / Heterogenität: Hierunter werden Unterschiede zwischen den Studienpopulationen, den Interventionen, den Settings und den erhobenen Outcomes verstanden. (Im Prinzip ist es dann legitim Studienergebnisse zu poolen, wenn die Studien sich so ähnlich sind wie die Arme eines Multicentertrials).

b) Sind die Überlegungen zu Punkt a) positiv bewertet worden, ist zu entscheiden, welches Meta-Analyse-Verfahren eingesetzt werden soll. Dies muss folgende Aspekte berücksichtigen:

- - Fragestellung und Studiendesign: Therapie?, Diagnostik?, Prognose?
- - Effektmaß (dichotome Ergebnisse, kontinuierliche Daten?) - Können die Ergebnisse aller Studien mit Hilfe eines Effektmaßes ausgedrückt werden?

Weitgehend standardisierte und erprobte Meta-Analyse-Verfahren liegen vor allem für therapeutische Studien, d. h. RCTs und andere kontrollierte Studien vor. Literatur und Computerprogramme geben bei der Auswahl und Durchführung der Berechnungen wertvolle Hilfestellung. Die schon erwähnte RevMan© Software der Cochrane Collaboration liefert auch gleich eine graphische Darstellung der Ergebnisse im "Forest Plot" und bietet die Möglichkeit zur graphischen Abschätzung des "publication bias". Auf eine kompetente methodische Beratung sollte jedoch keinesfalls verzichtet werden.

Die Methodik quantitativer Synthesen von Ergebnissen diagnostischer Studien oder observationeller Studien ist in einem weit geringeren Maße standardisiert und macht, falls im Rahmen eines HTA derartige Analysen durchgeführt werden sollen, die Konsultation von Methodikexperten zwingend erforderlich.

Es darf jedoch nicht vergessen werden, dass Meta-Analyse-Ergebnisse immer nur ein Aspekt der Ergebnisse eines HTA-Verfahrens sind, erst die Zusammenführung mit den Kontextinformationen stellt die Entscheidungsgrundlage für das Problem her. Informationen aus qualitativ hochwertigen Publikationen, die bereits Informationssynthesen enthalten, werden in der Regel keinem weiteren systematischen Syntheseverfahren unterzogen, sondern lediglich nach ihren Kernaussagen in die qualitative Informationssynthese eingebracht.

Die Einordnung der Ergebnisse der Informationssynthesen in den Zusammenhang mit Informationen zum Status Quo der Technologie, zur Epidemiologie der Zielerkrankung sowie zur Indikationsstellung erfolgt in einer bewertenden Diskussion, in der auch die weitergehenden Implikationen, wie soziale, rechtliche, gesundheitspolitische und ethische Aspekte, Beachtung finden.

Die Ergebnisdiskussion im Review bleibt also begrenzt auf solche Aspekte, die relevant für klinische Entscheidungen bzw. für die Identifikation von Forschungsbedarf sind. Hierzu gehören:

- Methodische Einschränkungen der Studienqualität, soweit sie die Aussagekraft des Reviews beeinträchtigen.

- Wie die eingeschlossenen Primärmaterialien im Kontext zu anderen wissenschaftlich belegten Aussagen einzuordnen sind (mit der Angabe, inwieweit die anderen Materialien systematisch ausgewertet wurden).
- Wie die Ergebnisse in die übliche klinische Praxis einzuordnen sind (unter Beachtung der Tatsache, dass auch die klinische Praxis erhebliche Variationen aufweist).
- Konsequenzen der Ergebnisse für unterschiedliche Interessengruppen im Gesundheitsbereich, einschließlich der Verbraucher.

Es fließen Qualitätskriterien der Einzelstudie sowie der entsprechenden Zusammenstellung mit ein: Grundlage dieses Vorgehens ist die Beobachtung, dass methodisch mangelhafte Studien dazu tendieren, das Ergebnis zu überschätzen. Daher müssen solche Studien unter Vorbehalt in den Review aufgenommen oder sogar ganz ausgeschlossen werden.

I C.3.5.2 Methoden: Modifikations- / Kürzungspotential

Wie oben bereits angedeutet, kann, da im Methodenmanual für die deutsche HTA-Arbeitsgruppe bisher keine Standards für die Informationssynthese erarbeitet wurden, auch keine Kontrastierung / Kürzung vorgenommen werden. Die hier ausgesprochenen Empfehlungen zum Prozedere im Rahmen des Kurz-HTA wurden aus international etablierten Arbeitsmethoden abgeleitet.

Sind die Studien bereits aufgrund ihrer klinischen Charakteristika extrem unterschiedlich, muss auf das Poolen von Daten (Meta-Analyse) verzichtet werden. Sofern die Datenlage es jedoch erlaubt, sollte eine gemeinsame statistische Auswertung der eingeschlossenen Studien durchgeführt werden. Hierbei empfiehlt sich die Beschränkung auf die wichtigsten Zielgrößen, die vorab im Protokoll festgelegt werden sollten. Ein sorgfältiges Vorgehen bei der statistischen Auswertung ist unerlässlich, da sonst verfälschte Resultate zu erwarten sind (Ziegler 1999). Bei der Auswertung sollte unter anderem potentielle Heterogenität zwischen den Studien gründlich untersucht werden, eine Reihe von Sensitivitätsanalysen durchgeführt werden und das Vorliegen von Publikationsbias untersucht werden. Als einzige Verkürzung dieses Arbeitsschrittes könnte unserer Meinung nach der Verzicht auf doppelte Darstellung der gleichen Informationen, nämlich quantitativ und qualitativ in Erwägung gezogen werden.

I.C.3.5.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt

Für das aktuelle Kurz-HTA-Projekt entsprachen ein HTA-Bericht sowie fünf prospektive Kohortenstudien den Einschlusskriterien.

Wir analysierten den HTA-Bericht ausführlich unter den Gesichtspunkten Bezugsrahmen, konkrete Fragestellung, verwendete Methodik, Ergebnisse und Schlussfolgerungen im fortlaufenden Text.

Die Charakteristika und Hauptergebnisse der fünf prospektiven Kohortenstudien wurden ebenfalls ausführlich einzeln in Textform vorgestellt, zusätzlich erfolgte jedoch noch eine Dokumentation in tabellarischer Form.

- Bei insgesamt fünf Studien war der Arbeitsaufwand, eine beschreibende im fortlaufenden Text stattfindende Informationsdarstellung anzubieten gering.
- Bei nur fünf Studien, die sich auch noch als sehr heterogen darstellten (unterschiedliche technische Geräte, unterschiedliche Ursprungsländer etc.), verbietet es sich, eine Meta-Analyse zu rechnen, stattdessen bietet sich eine beschreibende Textdarstellung an.

I C.3.5.4 Diskussion: Bewertung, eigene Erfahrungen

Die Zusammenfassung wie oben beschrieben, gestaltete sich problemlos.

I C.3.5.5 Schlussfolgerung: SOP 5

Vorschlag für eine allgemeingültige Arbeitsanleitung:

- 1. Sichtung** der nach SOP 4 extrahierten Informationen
- 2. Obligate tabellarische Darstellung** von Studienqualität, Studienpopulation, Intervention, Zielgrößen, Settings, Outcomes (Art, Größe, Robustheit)
- 3. Beschreibung der zentral relevanten Aspekte** in einem Begleittext; Cave: keine Wertungen, diese sind Gegenstand der Diskussion
- 4. Entscheidung**, ob für (einen) Einzelaspekt(e) ein Datenpooling erforderlich und machbar ist. Dazu bitte entscheiden:
 - Ist das Poolen von Ergebnissen erforderlich, um zu verwertbaren Aussagen zu kommen? Falls nein, weiter mit der Diskussion.
 - Sind die Studien klinisch und methodisch so homogen (Könnten sie Arme einer grossen Multicenterstudie sein?), dass ein Zusammenführen der Ergebnisse sinnvoll erscheint? Falls nein, weiter mit der Diskussion.
 - Welches Verfahren soll zum Poolen eingesetzt werden? Gegebenenfalls Rücksprache mit statistischer Abteilung, Durchführung der Meta-Analyse.

I C.3.6 SOP 6: Schlussfolgerungen, Aktualisierungsbedarf

I C.3.6.1 Hintergrund: Status Quo

Die Vorgaben zur Formulierung von Schlussfolgerungen im Rahmen eines HTA-Gutachtens durch das Methodenmanual der deutschen HTA-Arbeitsgruppe sind richtungsweisend, aber wenig präzise. Es wird angedeutet, dass die HTA-Gutachten entweder mit einer wertungsfreien Zusammenfassung der Evidenz, ohne Aussprache von Empfehlungen abschließen, oder aber explizite Empfehlungen für oder gegen den Einsatz von Verfahren, ihre Finanzierung, Auflagen oder Indikationskorridore aussprechen können. Die Identifizierung von Forschungsbedarf gehört ebenso zu den Schlussfolgerungen wie die Festlegung eines Termins für die zeitgemäße Aktualisierung und die Planung eines Effektmonitoring.

Auch den internationalen Methodenmanualen für HTA sind nur sehr begrenzt Informationen zur systematischen Ableitung von Schlussfolgerungen aus aufgefundener Evidenz zu entnehmen (Kristensen et al. 2001; Liberati et al. 1997; DEC 2000; Burls et al. 2000; NICE 2000; ECHTA/ECAHI Project 2001). In den meisten Manualen wird vor allem inhaltlich umrissen, welche Aspekte in den Schlussfolgerungen angesprochen werden sollten.

Die ECHTA/ECAHI-Arbeitsgruppe hält fest, dass die Schlussfolgerungen eines HTA-Gutachtens:

- Antworten auf die Forschungsfragen geben sollen,
- Qualität und Herkunft der zugrundeliegenden Evidenz darlegen,
- eine Zusammenfassung der Evidenz zu allen bewerteten Fragestellungen geben,
- Beschreibung der Effektgrößen (positiv und negativ) liefern,
- falls gruppenspezifische Unterschiede gefunden wurden: Beschreibung geben,
- falls technologieabhängige Unterschiede gefunden wurden: Beschreibung geben,
- Einordnung der Evidenz in den nationalen / lokalen Kontext vornehmen und
- Forschungsbedarf nennen.

Auch für die Ableitung von Empfehlungen gilt der Anspruch an Validität verbunden mit dem Anspruch an Transparenz. Systematische Arbeiten zur Ableitung von Empfehlungen / Schlussfolgerungen aus unterschiedlichen Evidenzgrundlagen kommen vor allem aus Arbeitsgruppen, die mit der Erstellung von Leitlinien befasst sind. Leitlinien müssen in der Regel zu allen Aspekten, die bei der Versorgung von Patienten mit einem bestimmten Störungsbild relevant sind, Stellung nehmen bzw. Empfehlungen abgeben. Für viele dieser Empfehlungen fehlt eine solide wissenschaftliche Evi-

denzbasis (oder ist in nur sehr beschränktem Umfang verfügbar), sodass die einzelnen Leitlinienempfehlungen auf Informationen sehr heterogener Qualität stehen. Um diesem Umstand Rechnung zu tragen, wurden von einigen Arbeitsgruppen sogenannte Grading Systeme entwickelt, die für den Leser auf den ersten Blick kenntlich machen sollen, wie die Evidenzgrundlage von Empfehlungen / Schlussfolgerungen beschaffen ist.

Die unseres Wissens ältesten Arbeiten entstanden im Zuge der Aktivitäten der "Canadian Task Force on the Periodic Health Examination" (CTFPHE 1979) und wurden für die Entwicklung der AHCPH Clinical Practice Guidelines weiter entwickelt (Field et al. 1990; US Department for Health and Human Services 1993). Das vorgeschlagene Verfahren zur Generierung von Empfehlungen bestand aus zwei Schritten: Im ersten Schritt werden die einer Empfehlung zugrunde liegenden Primärmaterialien aufgrund ihres Studiendesigns in eine Evidenzhierarchie eingeordnet, im zweiten Schritt wird der Empfehlung anhand der Qualität und Menge der zugrunde liegenden Studienergebnisse ein Evidenzgrad zugeordnet.

Das Grading-System der AHCPH in der durch das schottische Leitlinieninstitut SIGN (Scottish Intercollegiate Guideline Network) modifizierten Form ist in Tabelle 17 dargestellt.

Tabelle 17: Grading System von SIGN, modifiziert nach AHCPH (1993) Ü. d. A.

Klassifikation der Evidenzstufen	
Ia	Evidenz aus einer Meta-Analyse von randomisierten kontrollierten Studien
Ib	Evidenz aus wenigstens einer randomisierten kontrollierten Studie
IIa	Evidenz aus wenigstens einer methodisch gut durchgeführten kontrollierten Studie
IIb	Evidenz aus wenigstens einer methodisch gut durchgeführten quasi-experimentellen Studie
III	Evidenz aus methodisch gut durchgeführten nicht-experimentellen Studien (Beobachtungsstudien, Querschnittsstudien und Fallserien)
IV	Evidenz aus Expertengutachten und / oder -voten
Evidenzgrade der Empfehlungen	
A	Erfordert wenigstens eine randomisierte kontrollierte Studie als Teil einer insgesamt guten und konsistenten Literaturbasis (Evidenzstufen Ia und Ib)
B	Erfordert das Vorliegen von methodisch gut durchgeführten klinischen Studien aber keine RCTs (Evidenzstufen IIa, IIb, III)
C	Evidenz aus Expertengutachten und / oder -voten und signalisiert, dass derzeit keine klinischen Studienergebnisse verfügbar sind

Dieses System wird noch weit verbreitet eingesetzt und hat nicht zuletzt durch die AHCPH Praxisleitlinien einen gewissen Bekanntheitsgrad erlangt. Dennoch sind in Diskussionen zwei Hauptkritikpunkte laut geworden:

- Die Evidenzhierarchie bezieht sich streng auf den Bereich Therapie. Empfehlungen zu diagnostischen Verfahren oder prognostischen Aussagen lassen sich anhand dieses Systems nicht graduieren.
- Die Anwendung der Evidenzhierarchie setzt per se zunächst keine weitere Qualitätsstaffelung innerhalb der Kategorien voraus - z. B. würden die Ergebnisse einer methodisch fragwürdigen randomisierten kontrollierten Studie im Evidenzgrad gleichgesetzt mit methodisch einwandfreien RCTs.

Weiterentwicklungen der Graduierungssysteme von Hadorn et al. (1996) und vor allem vom Center of Evidence Based Medicine in Oxford (CEBM 2001) tragen vor allem dem ersten Kritikpunkt Rechnung. Hadorn et al. (1996) entwarfen eine universal einsetzbare Checkliste, die sich auch für die Bewertung von nicht-randomisierten Studien eignet. Am CEBM wurden separate Evidenzhierarchien mit den entsprechenden Empfehlungsgraduierungen entwickelt für die Problembereiche Therapie / Prävention / Schädigung, Diagnose, Prognose und ökonomische Analysen.

Die Weiterentwicklung der SIGN-Graduierung berücksichtigt vor allem den zweiten Gesichtspunkt. Die Anleitung zur Verwendung des Schemas beinhaltet die Anweisung, jede einzelne Primärquelle nicht nur einer Kategorie zuzuordnen, sondern auch eine ausführliche Qualitätsbewertung (Instrumente hierfür wurden für die Studientypen systematischer Review, RCT, Kohortenstudie, Fall-Kontrollstudie und diagnostische Studie entwickelt) vorzunehmen. Anhand dieser Qualitätsbewertung ist es nun möglich, dem Evidenzgrad ein ++ (qualitativ-methodisch hochwertig, very low risk of bias), + (qualitativ-methodisch akzeptabel, low risk of bias) oder ein - (qualitativ-methodisch inakzeptabel, high risk of bias) hinzuzufügen. Dies weiterentwickelte Graduierungsschema berücksichtigt ausserdem, ob Studienergebnisse auf die interessierende Studienpopulation angepasst (extrapoliert) werden mussten. Sind relevante Anpassungen erforderlich, wird die Graduierung eine Stufe tiefer ausfallen, als von der Evidenzbasis her abzuleiten wäre.

Auch die Evidenzhierarchie des CEBM berücksichtigt die Möglichkeit methodischer Mängel innerhalb einer Studiendesignkategorie. Bei den Therapiestudien erhalten "low quality RCTs" nur den Evidenzgrad 2b (und damit den Status von Kohortenstudien) anstelle von 1b.

Graduierte Empfehlungen sind vor allem dann sinnvoll, wenn in einem Dokument mehrere Empfehlungen, denen unterschiedlich valide Evidenz zugrunde liegt, ausgesprochen werden müssen. Dies wird in HTA-Verfahren jedoch eher die Ausnahme sein.

I C.3.6.2 Methoden: Modifizierungs- / Kürzungspotential

Generell sollte gelten: Die Schlussfolgerungen beantworten die im ersten Arbeitsschritt (SOP 1) abgestimmten wenigen präzisen Forschungsfragen. Ob nur Schlussfolgerungen gezogen oder explizite Empfehlungen ausgesprochen werden sollen, ist

im Zuge der Abstimmung der Fragestellungen besonders auch mit den (potentiellen) Nutzern zu klären.

Evidenzbasis und Beschreibung der Ergebnisse

Für die Kernfrage nach der Effektivität der Technologie hat

1. die Feststellung der Evidenzbasis und
2. eine geeignete Darstellung der Ergebnisse

zu erfolgen.

Ad 1: Eines der kennzeichnenden Elemente eines Kurz-HTA-Verfahrens ist die starke Einengung der zu bearbeitenden Fragestellung auf ein bis zwei Kernelemente. Eine aufwendige Graduierung von Empfehlungen kann daher entfallen, weil sich für diese überschaubare Anzahl von Aspekten die Evidenzbasis sehr unaufwändig im Text oder in einer kurzen Tabelle darstellen und kommentieren lässt. Dabei sollte eine Zuordnung der Evidenz zu einem Evidenzlevel vorgenommen werden. Wir empfehlen hier die Verwendung der Evidenzlevel nach CEBM (2001), weil sich mit ihrer Hilfe einerseits die Evidenzgrundlagen nicht nur für therapeutische, sondern auch für prognostische, diagnostische und sogar gesundheitsökonomische Fragestellungen abbilden lassen und andererseits die Möglichkeit einer Berücksichtigung der Studienqualität innerhalb der Kategorien gegeben ist. Weiterhin sollte ein Hinweis eingefügt werden, ob die Studienergebnisse auf die eigene Entscheidungssituation übertragbar sind oder ob ihre Umsetzung Extrapolationen erfordert. Dies fügt dem ermittelten Evidenzgrad eine negative Tendenz hinzu.

An dieser Stelle sollte, neben der Erwähnung von rein methodischen Schwächen oder Stärken der Evidenzbasis, auch auf die möglicherweise aus inhaltlichen Gründen notwendige Extrapolation von Studienergebnissen hingewiesen werden. Dabei wären vor allem Unterschiede zwischen interessierender und untersuchter Population, Intervention, Outcome, Setting oder Klasseneffekte (Mc Alister et al. 1999) mit ihren vermuteten Auswirkungen auf die Ergebnisinterpretation zu erwähnen.

Ad 2: Um die Ergebnisdarstellung zu vereinheitlichen, sollten geeignete, allgemein gebräuchliche Effektmaße verwendet werden. Zur Darstellung von Interventionsfolgen (erwünscht oder unerwünscht) und der Auswirkung von Expositionen, Risiko- oder Prognosefaktoren sollten regelhaft absolute und relative Effektmaße gebraucht werden. Im Einzelnen kommen, bei dichotomen bzw. dichotomisierten Zielgrößen, die in Tabelle 18 aufgeführten Effektmaße hierzu infrage.

Werden die Ergebnisse in nicht-dichotomisierter Form angeboten, ist der Ergebnisparameter der Wahl die Differenz der Mittelwerte.

Tabelle 18: Effektmaße zur Beschreibung von dichotomen Studienergebnissen

Absolute Effektmaße	
absolutes Risiko (AR)	Anteil von Personen einer Untersuchungsgruppe, bei dem die Zielgröße festgestellt wurde; $(n_{\text{Zielgröße}} / n_{\text{Gesamt}}) \times 100 = \text{absolutes Risiko in \%}$
absolute Risikoreduktion (ARR)	Differenz der absoluten Risiken zwischen zwei verschiedenen Untersuchungsgruppen (z.B. exponierte und nicht-exponierte Personen in einer Kohortenstudie $AR_{\text{exp.}} - AR_{\text{nicht-exp.}} = \text{ARR in \%}$)
attributables Risiko (absolut)	Risikoanteil am Gesamtrisiko, der auf eine Exposition zurückzuführen ist; ergibt sich: - aus der Differenz von Gesamtrisiko und Risiko in der nicht-exponierten Gruppe ($AR_{\text{Gesamt}} - AR_{\text{Nicht-exponiert}} = \text{auf Exposition zurück zu führender Risikoanteil}$)
Number Needed to Treat (NNT), Number Needed to Treat to Harm one patient (NNH)	Anzahl von Personen, die behandelt werden müssen (exponiert sein müssen), um ein zusätzliches Zielereignis zu erreichen / zu verhindern (je nach Fragestellung). Kehrwert der ARR ($1 / \text{ARR}$)
Relative Effektmaße	
relatives Risiko (RR)	Quotient aus den absoluten Risiken der zu vergleichenden Untersuchungsgruppen (z.B. aus RCT: $AR_{\text{Therapiegruppe}} / AR_{\text{Kontrollgruppe}} = \text{RR}$)
relative Risikoreduktion (RRR)	Beschreibt den Anteil am Gesamtrisiko der z. B. durch Einsatz einer Intervention beseitigt werden kann: $(AR_{\text{Kontrollgruppe}} - AR_{\text{Therapiegruppe}}) / AR_{\text{Kontrollgruppe}} = \text{RRR in \%}$
attributables Risiko (relativ)	Risikoanteil am Gesamtrisiko, der auf eine Exposition zurückzuführen ist; ergibt sich: aus der Differenz von Gesamtrisiko und Risiko in der nicht-exponierten Gruppe bezogen auf das Gesamtrisiko $(AR_{\text{Gesamt}} - AR_{\text{Nicht-exponiert}}) / AR_{\text{Gesamt}} = \text{Anteil am Gesamtrisiko, der auf Exposition zurück zu führen ist}$
Odds Ratio (OR)	"Chancenquotient" wird berechnet, wenn keine absoluten Risiken berechenbar sind (z.B. aus Fall-Kontrollstudien). ("Fälle" _{exponiert} / "Fälle" _{nicht-exponiert}) / (Kontrollen _{exponiert} / Kontrollen _{nicht-exponiert}); Odds Ratios werden auch berechnet, wenn bei der Datenanalyse in multivariaten Verfahren Störgrößen berücksichtigt werden müssen.

Die Ergebnisse von Evaluationsstudien zu diagnostischen Verfahren werden anhand von Parametern dargestellt, die beschreiben, wie gut ein Verfahren in der Lage ist, zwischen Gesunden und Kranken zu diskriminieren. Hierzu gehören in erster Linie (Tabelle 19):

Tabelle 19: Effektmaße zur Beschreibung der Ergebnisse diagnostischer Studien

Sensitivität	Anteil testpositiver Personen an allen Erkrankten in der Untersuchungskohorte
Spezifität	Anteil testnegativer Personen an allen Nicht-Erkrankten in der Untersuchungskohorte
Likelihood Ratio (eines positiven Testergebnisses*)	Beschreibt das Verhältnis der Wahrscheinlichkeit eines positiven Testes bei Kranken zur Wahrscheinlichkeit, dass ein positiver Test bei Gesunden vorliegt.
Positiv prädiktiver Wert*	Anteil erkrankter Personen an allen Personen mit positivem Testergebnis

* Die Parameter für negative Testergebnisse errechnen sich in Analogie.

Die Unsicherheit aller Ergebnisparameter wird durch die Angabe von Konfidenzintervallen beschrieben (ein Bereich berechneter Werte, der den wahren Wert einer Statistik mit angegebener Wahrscheinlichkeit überdeckt). Bei medizinischen Fragestellungen wird in der Regel das 95% Konfidenzintervall angegeben.

Unbedingt beachtet werden sollte allerdings, dass die oben genannten Ergebnismaße nur dann berichtet werden können, wenn sie in den Studien angegeben sind bzw. wenn die erforderlichen Rohdaten zu ihrer Berechnung angegeben sind. Eigene Schätzungen auf der Grundlage von Summationsgrößen gehören nicht zur Ergebnisberichterstattung und sollten, wenn überhaupt, in einem separaten Gliederungspunkt der Schlussfolgerungen vorgenommen werden.

Generierung von Empfehlungen / Schlussfolgerungen

Nach Darlegung der Evidenzbasis muß diese in Empfehlungen oder Schlussfolgerungen überführt werden, die, für HTA typisch, konkreten Entscheidungen zugrunde gelegt werden sollen. Der Entscheidungs- oder Problemdruck kann eher hoch, oder eher niedrig sein und aus unterschiedlichen Quellen resultieren (finanzieller Druck; Öffentlichkeit wie Presse, Publikum oder Politik; technischer Fortschritt; Epidemiologie; demographischer Wandel u. a.). Zusammenfassend lässt sich das Verhältnis von Entscheidungsdruck zur Evidenzlage durch vier Konstellationen beschreiben:

	Problem- / Entscheidungsdruck	
	eher hoch	eher niedrig
Evidenz stark	a	B
Evidenz schwach	c	d

Abbildung 1: Verhältnis von Entscheidungsdruck zur Evidenzbasis

Ohne Probleme werden sich starke (eindeutige) Empfehlungen für die Situationen in den Zellen a und d aussprechen lassen. Die Konstellationen in b und c dagegen werden mehr Argumentation, auch neben der reinen "Evidenzbewertung", erfordern. Zu den modifizierenden Faktoren, die hier (mehr als in Zellen a und d) anzuführen

sind, gehören beispielsweise: unerwünschte Wirkungen, Compliance (Patienten, Leistungserbringer), Infrastruktur, Qualitätssicherung, Kosten, "Moral Hazards" usw..

Forschungsbedarf

In den seltensten Fällen wird ein HTA-Verfahren alle gestellten Fragen zufriedenstellend anhand valider Evidenz beantworten können. Evidenzlücken können dabei sowohl den klinischen als auch den versorgungsrelevanten Kontext betreffen. Die Formulierung von Forschungsbedarf wird damit zum obligaten Bestandteil fast aller Schlussfolgerungen von HTA-Berichten.

Aktualisierungsbedarf

Die Feststellungen eines HTA-Berichtes stellen immer nur eine Momentaufnahme des Entwicklungs- und Einsatzprozesses eines Verfahrens dar. Je nach Geschwindigkeit dieser Prozesse werden auch die Ergebnisse von Verfahrensbewertungen schneller oder langsamer überholt und müssen aktualisiert werden. Es gibt eine Reihe von Faktoren, die die Geschwindigkeit von Veränderungen in der Versorgung und des gesundheitspolitischen Kontext erheblich beeinflussen - ihr Einfluss auf den Einsatz des bewerteten Verfahrens bestimmt, in welchen Abständen der HTA-Bericht zu aktualisieren ist. Zu den Faktoren gehören: Die demographische Entwicklung, die Weiterentwicklung der Technologie, die Evidenzlage (neue Studienergebnisse), Änderungen in Anspruchs- und Kostenentwicklung (Bedürfnisdynamik) oder unerwünschte soziokulturelle Auswirkungen.

Für das Kurz-HTA wäre unter Berücksichtigung dieser Faktoren eine Art "Verfallsdatum" zu benennen, nach dessen Ablauf die Schlussfolgerungen nicht mehr umgesetzt werden sollten.

I C.3.6.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt

Im Kurz-HTA zur Ultraschallmessung der Knochendichte wurde die Beantwortung der fünf Reviewfragestellungen in der Diskussion vorgenommen.

Feststellung der Evidenzbasis und zusammenfassender Ergebnisbericht erfolgten im Rahmen der Beantwortung der Kernfrage (Frage 1: Wie valide lässt sich das Frakturrisiko mit Hilfe der quantitativen Ultraschall-Verfahren im Vergleich zum DXA-Verfahren abschätzen?). Die Evidenzbasis umfasste einen HTA-Bericht mit systematischer Literaturübersicht über prospektive Kohortenstudien und fünf einzelne prospektive Kohortenstudien. Dem HTA-Bericht wurde die Evidenzklasse 4, den Einzelstudien jeweils die Evidenzklasse 1b nach CEBM 2001 zugeordnet. Der HTA-Bericht erhielt die niedrige Einstufung, weil ihm eine Dokumentation der Qualität der eingeschlossenen Primärmaterialien fehlte. Der Evidenzlevel 4 beschreibt somit die "worst case" Annahme (dass es sich bei den eingeschlossenen Primärstudien um methodisch mangelhafte Arbeiten handelt). Die Einzelstudien wurden mit 1b bewertet, mit-

hilfe unseres Bewertungsinstrumentes (Stufenplan 2) fanden wir alle zentralen Qualitätskriterien für prognostische Kohortenstudien erfüllt. Die in den Studien untersuchten Studienpopulationen entsprachen jedoch nur einer Subgruppe der uns interessierenden Gesamtpopulation, sodass die Übertragbarkeit nicht ohne (diskutierte) Einschränkungen möglich war.

Zur Ergebnisdarstellung wurden in den Studien relative Effektmaße angegeben (relative Risiken, Odds Ratios), in einer Studie auch diagnostische Sensitivitäten zur Erkennung von frakturgefährdeten Personen.

Das Verhältnis von Entscheidungs- / Problemdruck zur Evidenzbasis entspricht in etwa der Zelle d der in Abbildung 1 gezeigten Vierfeldertafel. Der Entscheidungs- / Problemdruck ist zur Zeit noch nicht all zu hoch (es steht keine Entscheidung zur Aufnahme in den GKV-Leistungskatalog an; das Verfahren ist sicher, führt bei untersuchten Personen zu keinen physischen Schädigungen; die öffentliche Nachfrage hält sich (noch) in Grenzen, die Evidenzlage für seinen prognostischen Nutzen muss als mäßig beurteilt werden. Trotz in sich valider Studien (Klasse 1b) werden deren (eher schwache) Aussagen durch eingeschränkte Übertragbarkeit weiterhin relativiert.

Diese Konstellation bewirkt für die Schlussfolgerungen des Kurz-HTAs, dass eine recht ausführliche Diskussion neben der reinen wissenschaftlichen Evidenz auch andere Aspekte des Verfahrenseinsatzes heranzieht (wie in den Reviewfragestellungen vorgesehen), und dass eine kurze "Gültigkeitsdauer" des Kurz-HTA veranschlagt wird (2 Jahre nach Redaktionsschluss).

I C.3.6.4 Diskussion: Bewertung, eigene Erfahrungen

Schwierigkeiten bei der Umsetzung der Vorgaben für "Schlussfolgerungen und Empfehlungen" zeigte sich vor allem an einer Stelle: Die Zusammenführung der Informationen zu interner und externer Validität der vorliegenden Studien (HTA-Bericht und Einzelstudien) war nicht einfach. Eine Bewertung der internen Validität (Einstufung in Hierarchie) wird vor dem Hintergrund zu überdenken sein, dass die "Stufenpläne" nur den Einschluss von Materialien hoher interner Evidenzklasse zulassen - andere werden "ausgefiltert". Andererseits gibt das hier vorgeschlagene SOP noch relativ ungenaue Anweisungen zum Umgang mit Problemen der externen Evidenz.

I C.3.6.5 Schlussfolgerung: SOP 6

Vorschlag für eine allgemeingültige Arbeitsanleitung

Die hier als Schlussfolgerungen bezeichneten Ausführungen können im eigentlichen HTA-Bericht je nach Umfang in einer "Diskussion" oder im Gliederungspunkt "Schlussfolgerungen" angesprochen werden.

1. Beantwortung der Berichtsfragestellungen

Die Schlussfolgerungen beginnen mit der Beantwortung der Berichtsfragestellung. Als erstes wird die Kernfrage zur Effektivität anhand der vorliegenden Evidenz beantwortet. Hierzu erfolgt:

- Die Benennung der Evidenz nach Menge und Qualität und die Zuordnung zu Evidenzlevels in der Hierarchie nach CEBM (2001). Einschätzung der Übertragbarkeit bzw. des Extrapolationsbedarfs (falls vorhanden, Korrektur des Evidenzlevels nach unten).
- Die Beschreibung der Studienergebnisse unter Verwendung der gängigen absoluten (AR, ARR, attributables Risiko) und relativen (RR, OR) Ergebnismaße. [Bei diagnostischen Studien Sensitivität, Spezifität und prädiktive Werte].
- Die Benennung von Extrapolationsbedarf - gibt es Diskrepanzen zu den in den Studien beschriebenen Studienpopulationen, Verfahren, Zielgrößen oder Settings - wie wirken sich diese auf die Ergebnisinterpretation aus?
- Beantwortung der peripheren Berichtsfragestellungen.

2. Generierung von Empfehlungen - falls vereinbart

Wie stellt sich das Verhältnis von Entscheidungsdruck / Evidenzbasis dar (vergl. Abbildung 1):

a: hoch / stark b: niedrig / stark c: hoch / schwach d: niedrig / schwach ?

Im Falle von a und d Konstellationen kann die Ableitung von Empfehlungen relativ einfach erfolgen; im Falle von b und c Konstellationen sind Kontextfaktoren wie unerwünschte Wirkungen, Compliance, Infrastruktur usw. zu berücksichtigen.

3. Formulierung von Forschungsbedarf

Zur Deckung von Evidenzlücken, die sowohl den klinischen (Patienten, Intervention, Outcome, Setting) als auch den versorgungsrelevanten Kontext (Epidemiologie, Infrastruktur, Qualität, Zugänglichkeit, Ökonomie, soziokulturelle oder rechtliche Aspekte) betreffen können.

4. Festschreiben eines "Verfalldatums"

Einschätzung der Technologieentwicklung, des Publikationsverhaltens (z. B. Informationen zu laufenden Studien?), der "Bedürfnisdynamik" und des Einflusses der demographischen Entwicklung. Aus diesen Informationen Vorschlag für die Gültigkeitsdauer.

I C.3.7 SOP 7: Dokumentationspflicht

I C.3.7.1 Hintergrund: Status Quo

Zentrales Qualitätskriterium von HTA-Verfahren ist die absolute Transparenz in allen Arbeitsschritten. Insbesondere bei den hier vorgeschlagenen Vorgehensweisen zur Erstellung eines Kurzgutachtens, die an mehreren Stellen Kürzungen und Abstriche bei der Reviewmethodik vornehmen, ist die sorgfältige, nachvollziehbare Dokumentation aller Schritte von zentraler Bedeutung, um mögliche Ansatzpunkte für systematische Verzerrungen der Reviewergebnisse aufzufinden.

Das Toolkit der German Scientific Working Group enthält Anleitungen und Formulare zur Dokumentation folgender Arbeitsschritte:

Tabelle 20: Dokumentationsvorschriften nach "German Scientific Working Group"

Arbeitsschritt	Dokumentationsbedarf
Literaturrecherche (vergl. auch SOP 2)	elektronische Datenbanken und Recherchestrategien mit Treffern; weitere Informationsquellen (Dokus 1-3)
Literaturbewertung (vergl. auch SOP 3)	Checklisten 1a (Kontextdokumente), 1b (systematische Reviews), 2a (Primärstudien, ausser Diagnosestudien), 2b (Diagnosestudien)

I C.3.7.2 Methoden: Modifizierungs- / Kürzungspotential

Wie oben angedeutet wird die Dokumentation für ein Kurz-HTA eher aufwändiger als für ein Standardverfahren ausfallen. Wir haben uns daher entschlossen, zwei separate Dokumentationsysteme einzuplanen. Die zentrale Dokumentation, die die Validitätsprüfung des Kurz-HTA-Berichtes zulässt, wird im Anhang des Berichtes angeführt. Eine Detaildokumentation kann bei besonderem Interesse von den Autoren angefordert werden.

Tabelle 21: Dokumentationsvorschläge für Rapid-HTA

SOP	Zentraldokumentation (Anhang des HTA-Berichtes)	Detaildokumentation (von den Autoren anzufordern)
SOP 1 (Fragestellung)	Zusammensetzung des Expertenkreises (Name, Adresse, fachliche Expertise) Modifiziertes Delphi-Verfahren: Fragestellungen vorher - nachher	Protokoll des modifizierten Delphi-Verfahrens: Korrespondenz, anonymisierte Stellungnahmen und Kommentare; zeitlicher Ablauf
SOP 2 (Literaturrecherche)	Ein- und Ausschlusskriterien; Datenbanken und andere Informationsquellen (Handsuche, Administration etc.); elektronische Suchstrategien und Treffer Vorgehen bei der manuellen Selektion einzuschliessender Publikationen (Liste)	ausgeschlossene Publikationen mit Begründung
SOP 3 (Qualitätsbewertung)	Tabellarische Darstellung der Ergebnisse der Qualitätsüberprüfung anhand der Stufenpläne	ausgefüllte Stufenpläne
SOP 4 (Datenextraktion)	ausgefülltes Datenextraktionsformular*	ausgefülltes Datenextraktionsformular*
SOP 5 (Informationssynthese)	tabellarische Darstellung der Studieneckdaten (evtl. identisch mit Dokumentation für SOP 4)	
SOP 6 (Schlussfolgerungen)	Einordnung der Ergebnisse in Evidenzhierarchie	

* kann je nach Umfang in den Bericht integriert sein oder ist auf Anfrage erhältlich

I C.3.7.3 Ergebnisse: Umsetzung im Kurz-HTA-Projekt

Die Dokumentationsvorschläge wurden umgesetzt (s. Teil II dieses Berichtes).

I C.3.7.4 Diskussion: Bewertung, eigene Erfahrungen

Die Dokumentationsvorschläge waren im Kurz-HTA problemlos umsetzbar.

I C.3.7.5 Schlussfolgerung: SOP 7

Vorschlag für eine allgemeingültige Arbeitsanleitung

Die in Tabelle 21 formulierten Dokumentationsvorschläge sollten übersichtlich, wenn möglich tabellarisch erfolgen. Für die Dokumentation innerhalb des Berichtes schlagen wir folgende Vorgehensweisen und / oder Musterformulare vor:

SOP 1	Expertenkreis	im Text / Anhang dokumentieren
	modifiziertes Delphi, Fragen	im Text / Anhang dokumentieren
SOP 2	Ein- / Ausschlusskriterien	im Text / Anhang dokumentieren
	Datenbanken:	

Datenbankdokumentationsformular

Tabelle 22: Dokumentationsformular: Datenbank

Datenbankname	Zugriff	Recherchierte Jahrgänge

Elektronische Suchstrategien und Treffer:

Tabelle 23: Dokumentationsformular: Recherche

Nr.	Treffer	Suchbegriff
# 1		
# 2		
# 3		

SOP 3 Ergebnisse der Qualitätsbewertung

Tabelle 24: Dokumentationsformular: Qualitätsbewertung

Quelle	Verwendeter Stufenplan		Punkte	Ein-/Ausschluss		Kommentar:
	1	2		ein	aus	
	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	
	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	

SOP 4	Datenextraktionsformular	tabellarisch, in Abhängigkeit von der Fragestellung anzupassen
SOP 5	Studieneckdaten	tabellarisch, in Abhängigkeit von der Fragestellung anzupassen
SOP 6	Evidenzgrade nach CEBM (2001) im Text / Anhang dokumentieren	

ID Literatur

1. Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ* 323:224-8; 2001
2. Antes G, Bassler D, Galandi D. Systematische Übersichtsarbeiten. Ihre Rolle in einer Evidenz-basierten Gesundheitsversorgung. *Dt Ärztebl* 96, 10:A-616-A-622; 1999
3. Banta HD, Luce BR. Health care technology and its assessment. An international perspective. Oxford University Press; 1993
4. Begg CB, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 276:637-9; 1996
5. Berlin JA, Drummond R. Measuring the Quality of Trials: The Quality of Quality Scales. *JAMA* 282(11); 1999
6. Berry E, Kelly St, Hutton J, Harris KM, Smith MA. Identifying studies for systematic Reviews. *Int J Tech Ass in Health Care* 16(2):668-72; 2000
7. Bitzer E, Busse R, Dörning H. Bestandsaufnahme, Bewertung und Vorbereitung der Implementation einer Datensammlung "Evaluation medizinischer Verfahren und Technologie" in der Bundesrepublik. 1. Auflage ed. Baden-Baden: Nomos Verlagsgesellschaft; 1998
8. Bucher HC. Kritische Beurteilung einer Übersichtsarbeit. *Münch med Wschr* 140:64-8; 1998
9. Burls A, Cummins C, Fry-Smith A, Gold L. West Midlands Development and Evaluation Service (DES), Department of Public Health and Epidemiology, University of Birmingham, UK, DPHE Report No 8; 2000
10. Chalmers I AD, editor. Systematic Reviews. London: Chalmers I, Altman DG; 2001
11. Canadian Task Force on the Periodic Health Examination: The Periodic Health Examination. *CMAJ* 121:1193-1253; 1979
12. Cochrane Reviewers' Handbook 4.1.3.
www.cochrane.dk/cochrane/handbook/handbook.htm updated June 2001; 2001
13. Undertaking Systematic Reviews of Research on Effectiveness.
www.york.ac.uk/inst/crd/report4.htm CRD Report Number 4(2nd Edition); 2001
14. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epidemiol* 48:167-71 ; 1995
15. DEC. DEC Guidelines. Wessex Institute for Health Research and Development; 2000
16. Deeks JJ. Systematic Reviews of evaluations of diagnostic and screening tests. *BMJ* 323:157-62; 2001
17. Droste S. Systematische Literaturrecherchen für die Erstellung von HTA-Berichten. *Z.ärztl.Fortbild.Qual.sich.(ZaeFQ)* 95:549-53; 2001

18. ECHTA/ECAHI Project: Working Group 4 Report: Best Practice in Undertaking and Reporting HTA. The European Commission Grant Agreement No. SI2.122594 (99CVF3-508); 2001
19. Egger M, Zellweger-Zähner T, Schneider M. Language bias in randomised controlled trials published in English and German. *Lancet* 350:326-9; 1997
20. Field MJ, Lohr KN (eds.) Institute of Medicine Committee to Advise the Public Health Service on Clinical Practice Guidelines. *Clinical Practice Guidelines: Directions for a new Programme*. Washington DC: National Academic Press; 1990
21. Flynn K. *Assessing Diagnostic Technologies*. Report No. 1. Office of Research and Development, VA Medical Center, Boston MA; MTA94-001-01; 1996
22. Garnero P, Dargent-Molina P, Hans D, Schott AM, Bréart, Meunier PJ, Delmas PD. Do Markers of Bone Resorption Add to Bone Mineral Density and Ultrasonographic Heel Measurement for the Prediction of Hip Fracture in Elderly Women? The EPIDOS Prospective Study. *Osteoporos Int* 8:563-569; 1998
23. Gibis B, Gawlik C, Sander G, Rheinberger P. Methodische Standards bei der Bewertung medizinischer Verfahren - Bericht über den Workshop der KBV am 23.5.2001 in Berlin. *Z.ärztl.Fortbild.Qual.sich.(ZaeFQ)* 95:584-6; 2001
24. Glüer CC, Barkmann R, Heller M. Quantitativer Ultraschall - Status 1999. *Radiologe* 39:213-221; 1999
25. Gnudi S, Ripamonti C, Malavolta N. Quantitative Ultrasound and Bone Densitometry to Evaluate the Risk of Nonspine Fractures: A Prospective Study. *Osteoporos Int* 11:518-523; 2000
26. Goodman CS, Ahn R. Methodological approaches of health technology assessment. *Int J Med Inf* 56:97-105; 1999
27. Grégoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: Is there a tower of Babel bias? *J Clin Epidemiol* 48:169-73; 1995
28. Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the Quality of Evidence for Clinical Practice Guidelines. *J Clin Epidemiol* 49(7):749-54 ; 1996
29. Hailey D, Corabian P, Harstall.C., Schneider W. The use and impact of rapid health technology assessments. *Int J Tech Ass in Health Care* 16(2):651-6; 2000
30. Hailey D, Topfer LA, Chan L, Wills F, Howell T. Horizon scanning for health technologies relevant to Alberta. Report on a pilot project. Alberta Heritage Foundation for Medical Research. Mai 2001
31. Hart D. Health Technology Assessment (HTA) und gesundheitsrechtliche Regulierung. *MedR* 1:1-8; 2001
32. Homik J, Hailey D. Quantitative ultrasound for bone density measurement. Alberta Heritage Foundation for Medical Research. HTA 11; 1998
33. Huang C, Ross PD, Yates AJ, Walker RE, Imose K, Emi K, Wasnich RD. Prediction of Fracture Risk by Radiographic Absorptiometry and Quantitative Ultrasound: A Prospective Study. *Calcif Tissue Int* 63:380-384; 1998

34. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan D, McQuay HJ. Assessing the Quality of Reports of Randomized Clinical Trials: Is Blinding Necessary? *Control Clin Trials* 17:1-12; 1996
35. Jüni P, Altman G, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 323:42-6; 2001
36. Jüni P, Witschi A, Bloch R, Egger M. The Hazards of Scoring the Quality of Clinical Trials for Meta-analysis. *JAMA* 282(11):1054-59; 1999
37. Kent DL, Larson EB. Magnetic resonance imaging of the brain and spine. Is clinical efficacy established after the first decade? *Annals of Internal Medicine* 108:402-24; 1988
38. Khan Kh.S., ter Riet G., Glanville J, Sowden AJ, Kleijnen J (eds.): *Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for those Carrying Out or Commissioning Reviews.* NHS Centre for Reviews and Dissemination, University of York, Report 4 (2nd Edition); 2001
39. Köbberling J; Richter K; Trampisch H-J, et al.. *Methodologie der medizinischen Diagnostik.* Berlin-Heidelberg: Springer Verlag; 1991
40. Koch A, Ziegler S. Metaanalyse als Werkzeug zum Erkenntnisgewinn. *Med Klin* 95:109-16; 2000
41. Kristensen FB, Horder M., Poulsen PB (eds.): *Health Technology Assessment Handbook.* Danish Institute for Health Technology Assessment; 2001
42. Langham J, Thompson E, Rowan K. Identification of randomized controlled trials from the emergency medicine literature: Comparison of hand searching versus MEDLINE searching. *Ann Emerg Med* 34:25-34; 1999
43. Laupacis A. Users' guides to the medical literature. IV. How to use an article about prognosis. *JAMA* 272(3):234-7; 1994
44. Liberati A, Sheldon TA, Banta D. EUR-ASSESS Project Subgroup on Methodology. Methodological Guidance for the conduct of Health Technology Assessments. *Int J Technol Assess Health Care* 13(2): 186-219; 1997
45. McAlister F, Laupacis A, Wells GA, Sackett DL. User's guide to the Medical Literature: XIX. Applying Clinical Trial Results. B. Guidelines for Determining Whether a Drug is Exerting more than a Class Effect. *JAMA* 282(14):1371-1377; 1999
46. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 356:1228-31; 2000
47. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, et al.. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 337:867-72; 2001
48. Moher D, Kadad A, Nichol G, Penman M, Tugwell P. *Assessing the Quality of Randomized Controlled Trials: An annotated Bibliography of Scales and Checklists.* *Control Clin Trials* 16:62-73; 1995

49. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 352:609-13; 1998
50. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 53:964-72; 2000
51. NICE: Guide to the Technology Appraisal Process. National Institute for Clinical Excellence UK; 2000
52. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-based Medicine Working Group. *JAMA* 272; 1994
53. Oxman AD, Guyatt GH, Singer J, Goldsmith Ch, Hutchison BG, Milner RA, Streiner DL. Agreement among Reviewers of Review Articles. *J Clin Epidemiol* 44(1):91-8; 1991
54. Perleth M, Lühmann D, Gibis B, Droste S. "Rapid Assessments" - schnelle Bewertung medizinischer Technologien. *Gesundheitswesen* 2001, 63(1):79-84; 2001
55. Perleth M, Wild C. Health Technology Assessment - eine Einführung. *TA-Datenbank-Nachrichten* Nr.1(10):4-12; 2001
56. Perleth M. Kritische Bewertung von HTA-Berichten. *Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis*. ed. Köln: Kunz R., Ollenschläger G, Raspe H, Jonitz G, Kolkman FW; 2000
57. Prins SH, Jorgensen HL, Jorgensen LV, Hassager C. The role of quantitative ultrasound in the assessment of bone: a review. *Clin Physiol* 18(1)3-17; 1998
58. Sackett DL. On the need for evidence-based medicine. *Evidence-Based Medicine* 1:5-6; 1995
59. Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 53:1119-29; 2000
60. Stevens A, Colin-Jones D, Gabbay J. Quick and clean: authoritative health technology assessment for local health care contracting. *Health Trends* 27(2):37-42; 1995
61. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-Analysis of Observational Studies in Epidemiology. A proposal for Reporting. *JAMA* 283(15):2008-12; 2000
62. US Department for Health and Human Services. Agency for Health Care Policy and Research. Acute Pain Management: Operative or medical procedures and trauma. Rockville MD, Clinical Practice Guideline No.1, AHCPR Publication No. 92-0023; 1993
63. Ziegler S, Lühmann D, Raspe H, Windeler J. Entscheidungsunterstützung durch Kurzstellungnahmen zur Evidenzlage (Verzicht auf) Anforderungen. *Z.ärztl.Fortbild.Qual.sich.(ZaeFQ)* 95:105-11; 2001
64. Ziegler S, Victor N. Gefahren der Standardmethoden für Meta-Analysen bei Vorliegen von Heterogenität. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 30:131-40; 1999

Exemplarisches "Kurz-HTA": Die Rolle der quantitativen Ultraschallverfahren zur Ermittlung des Risikos für osteoporotische Frakturen

II C Hauptdokument

II C.1 Policy Question / Ziel des HTA-Berichtes

Die am weitesten verbreitete Definition der Osteoporose geht auf eine Konsensuskonferenz von 1991 zurück: "Eine Erkrankung, die durch niedrige Knochenmasse und eine Beeinträchtigung der Mikroarchitektur des Knochengewebes gekennzeichnet ist und so zu einer erhöhten Knochenbrüchigkeit und daraus folgend zu einem erhöhten Risiko führt, Frakturen zu erleiden". Osteoporotische Frakturen und ihre Folgen bestimmen die durch die Osteoporose verursachte Krankheitslast. Eine effektive sekundärpräventive Strategie müsste somit die Identifikation von frakturgefährdeten Personen ermöglichen, um diesen anschließend eine wirksame und akzeptable Therapie anzubieten. Bisher war die osteodensitometrische Messung der Knochenmineraldichte (KMD) mit Röntgenstrahlen das einzige nicht-invasive Verfahren zur Abschätzung der Knochenbrüchigkeit und damit des intrinsischen Frakturrisikos. Die häufigsten angewendeten Methoden zur Messung der KMD am peripheren wie auch am Achsenskelett sind die Zwei-Spektren-Röntgenabsorptiometrie (Dual X-ray Absorptiometry = DXA) und die quantitative Computertomographie (QCT + pQCT).

Eine wesentliche Effektivität der derzeit gebräuchlichen osteodensitometrischen Verfahren zur Früherkennung einer Osteopenie/Osteoporose und damit Frakturgefährdung ist bisher nicht belegt. Schätzungen verschiedener Arbeitsgruppen ergaben, dass weniger als 10 % aller zu erwartenden Frakturen durch ein derartiges Programm verhindert werden könnten.

Vor dem Hintergrund fehlender ionisierender Strahlung und vergleichsweise geringerer Kosten haben in den letzten Jahren unterschiedliche quantitative Ultraschall-Verfahren (QUS) zunehmende Verbreitung bei der Abschätzung des Risikos für „osteoporotische“ Frakturen gefunden. Sie basieren auf der parametrischen Analyse der Wechselwirkung von mechanischen Schallwellen beim Durchgang durch den Knochen, wobei das Ultraschallsignal in Transmission aufgenommen wird. Zwei verschiedene Ultraschallbasismessgrößen werden hierbei untersucht: Die Schallgeschwindigkeit (ultrasound transmission velocity = UTV oder speed of sound = SOS) und die Breitbandultraschallabschwächung (broadband ultrasound attenuation = BUA). Ultraschallwellen mit unterschiedlichen Frequenzen werden, je nach vorhandener Dichte und Struktur der durchdrungenen Gewebe, abgeschwächt.

BUA und SOS werden nicht nur durch die Knochendichte, sondern wahrscheinlich auch durch Knochenstruktur und Knochenzusammensetzung beeinflusst (Barkmann 1999).

In Abhängigkeit von Messort (z. B. Kalkaneus, Fingerphalangen, Radius und – mit Einschränkungen – Tibia, Metatarsale und Patella) und Gerätetyp werden die Messergebnisse als Kombinationsparameter definiert und berichtet. An den Fingerphalangen wird z. B. die sogenannte amplitudenabhängige Schallabschwächung (amplitude dependent speed of sound = ADSOS) gemessen, für Kalkaneusmessungen dagegen werden gewichtete Mittelwerte aus BUA und SOS unter der Bezeichnung „stiffness“ oder „quantitative ultrasound index“ (QUI) ermittelt.

Eine besondere Attraktivität gewinnt die Methode durch die, im Gegensatz zu den radiologischen Verfahren, fehlende Strahlenexposition. Die QUS-Verfahren sind bislang nicht Bestandteil des EBM, werden jedoch zunehmend als IGeL-Leistungen in der ärztlichen Praxis und in Apotheken angeboten – die Entscheidung, eine Untersuchung durchführen zu lassen (und zu bezahlen), liegt damit bei den Patienten.

Die folgende Hintergrunddarstellung gibt zunächst einen Überblick über die angesprochene Zielerkrankung, beschreibt in der Folge das Verfahren unter technischen Aspekten und geht danach auf die Methodik zur Evaluation von Ultraschallgeräten ein.

II C.2 Hintergrund

II C.2.1 Hintergrund/Beschreibung des Krankheitsbildes

II C.2.1.1 Definitionen und Begriffsbestimmungen

Die Osteoporose ist eine systemische Skeletterkrankung, die durch niedrige Knochenmasse und eine Verschlechterung der Mikroarchitektur des Knochengewebes mit daraus folgender Zunahme der Knochenbrüchigkeit und gesteigertem Frakturrisiko charakterisiert ist. Das heißt: Die erhöhte Anfälligkeit der Knochen für Frakturen – also die Fragilität der Knochen – ist als die eigentliche Erkrankung anzusehen, mit den Folgen Frakturen und Deformationen des Skeletts. Einerseits wird die erhöhte Knochenbrüchigkeit (=Fragilität) als *Krankheit* beschrieben, andererseits ist sie *Risikofaktor* für das Erleiden von Knochenbrüchen. Die Entstehung von derartigen Frakturen wird als die Einwirkung der Kraft eines Sturzes aus nicht mehr als Stehhöhe auf ein fragiles Skelett beschrieben.

Die Fragilität von Knochen wird von mehreren Faktoren beeinflusst: Von der Knochenmasse, von histomorphometrischen Parametern und von den geometrischen Verhältnissen.

1. Unter diesen Faktoren läßt sich bisher nur die Knochenmasse über den Umweg des indirekt gemessenen Mineralgehaltes des Knochengewebes in vivo mit nicht-invasiven Verfahren quantitativ abschätzen. Die derzeit gebräuchlichen Verfahren sind vor allem die "Dual X-ray Absorptiometry" (DXA) und die "Quantitative Computertomography" (QCT an der Wirbelsäule; pQCT an Radius oder Tibia). In einigen wissenschaftlichen Arbeiten werden noch Messungen mittels Single Photon Absorptiometry (SPA), Single X-ray Absorptiometry (SXA), Dual Photon Absorptiometry (DPA) und peripheral Dual X-ray Absorptiometry (pDXA) berichtet, für die klinische Praxis sind diese Verfahren allerdings kaum noch relevant.

Zur Beschreibung der Messgröße wird im deutschen Sprachgebrauch zumeist der Begriff „Knochendichte“ verwendet, obwohl im physikalischen Sinne nur bei dreidimensional messenden Verfahren (QCT + pQCT) von Dichte als Einheit (Masse/Volumen, g/ml) gesprochen werden kann. Auch die Ergebnisse von zweidimensional messenden Verfahren (SPA, DPA, SXA und DXA) werden als Knochendichte bezeichnet, aber in der Einheit g/cm² angegeben.

2. Merkmale der intrinsischen Knochenqualität wie Vernetzung von Kollagenfibrillen oder das vermehrte Auftreten von Mikrofrakturen lassen sich derzeit in vivo und nicht-invasiv nicht erfassen. Das gleiche gilt für Merkmale der strukturellen Knochenqualität (z. B. Trabekelstruktur). Diese Parameter sind nur über eine Analyse von Knochenbiopsien zugänglich.

3. Mit Unterschieden in der Knochengometrie (z.B. Länge der Schenkelhalsachse oder Fläche der vertebralen Endplatten) lassen sich nach Heaney et al. (1998) und Ross et al. (1995) vermutlich unterschiedliche Frakturhäufigkeiten in den verschiedenen ethnischen Gruppen begründen. Experimente mit Leichenknochen weisen einen mäßigen bis starken Zusammenhang zwischen der Belastbarkeit sowohl von Femurknochen als auch von Wirbelkörpern mit der nicht-invasiv gemessenen Knochendichte nach. In-vitro-Studien zeigen, dass bis zu 80 % der Variabilität der Belastbarkeit von Femurknochen durch Dichtemessungen erklärbar sind. Die Variabilität der ex-vivo-Belastbarkeit (durch Kompression) von Wirbelkörpern unter experimentellen Bedingungen ist zu ca. 80 % aus der gemessenen Knochendichte erklärbar (Bouxsein et al. 1995).

Weil die Knochendichte zur Zeit der einzige nicht-invasive Parameter ist, der für die Beurteilung der Fragilität herangezogen werden kann, gelten die gemessenen Knochendichtewerte somit als Surrogat für Bruchfestigkeit bzw. Fragilität.

Die Entwicklung der Knochendichte unterliegt einem physiologischen Verlauf. Sie nimmt in der Kindheit und Adoleszenz zu und erreicht im jungen Erwachsenenalter einen Höhepunkt (peak bone mass). Die Höhe dieses Wertes wird von einer Reihe von verhaltens- und belastungsabhängigen Faktoren beeinflusst wie Körpergewicht, Kalziumgehalt der Ernährung, körperliche Aktivität und Rauchen (Earnshaw 1996). Nach Erreichen des Spitzenwertes nimmt die Knochendichte im Laufe des Lebens wieder ab (Davis 1989; Harris 1992). Deutlich nachweisbar sind diese Verluste vor allem an den stoffwechselaktiveren spongiösen Knochenarealen. Während der ersten 5 - 10 Jahre nach der Menopause kommt es bei Frauen zu einem deutlich ausgeprägten Mineralverlust: 1 – 2 % am peripheren Skelett, 4,5 % in der Wirbelsäule (Ettinger et al. 1985).

Die Verteilung der Knochendichtewerte entspricht in jedem Lebensalter einer Gauschen Normalverteilung (Lühmann et al. 2000). Die nach unterschiedlichen Methoden durchgeführten Knochendichtemessungen weisen eine erhebliche verfahrensabhängige Variabilität auf, sodass ein direkter Vergleich von Messwerten bzw. das Festlegen von „Normwerten“ erschwert ist. Diesem Umstand wurde Rechnung getragen, indem die Messwerte relativ als Einheiten von Standardabweichungen von Mittelwerten angegeben werden. Hierbei unterscheiden wir zwei Begriffe: Den T-Score, der sich auf den Mittelwert einer Referenzpopulation junger gesunder Frauen bezieht und den Z-Score, der sich auf den altersabhängigen Mittelwert bezieht. Der T-Score wird als Referenzwert empfohlen, da hier die Unterscheidung zur Knochendichte eines gesunden und stabilen Knochens herangezogen wird – somit also das Frakturrisiko am ehesten abgeschätzt werden kann (National Osteoporosis Foundation 1998).

II C.2.1.2 Häufigkeit von inzidenten oder prävalenten Skelettfrakturen/-deformitäten, Assoziation mit Osteoporose

In der populationsbezogenen prospektiven Erhebung von Wildner et al. (1996) wurde ein Algorithmus entwickelt, der die Unterscheidung zwischen „*intrinsischen*“ (d. h. eher durch altersassoziierte Pathomechanismen – wie abnehmende Knochendichte und zunehmende Sturzneigung – bedingte) von „*extrinsischen*“ (d. h. eher auf äußerliche Einwirkungen zurückzuführende) Frakturen ermöglicht und so die Ermittlung des vermutlich durch Alterungsprozesse bedingten Anteils an Frakturen erlaubt. Dieser altersassoziierte Anteil der Frakturen beinhaltet auch die osteoporoseassoziierten Frakturen, kann diesen allerdings nicht gleichgesetzt werden, da insbesondere im sehr hohen Alter auch andere Risikofaktoren für Frakturen eine zunehmend größere Rolle spielen. Häufig beobachtete Frakturlokalisationen sind Radius-, Wirbelkörper- und Oberschenkelhalsfrakturen, wobei letztere von hoher sozioökonomischer Bedeutung sind.

Da die Zahlen anhand von Arztkontakten ermittelt wurden, sind klinisch unauffällige Frakturen, also auch ca. zwei Drittel der Wirbelkörperfrakturen, nicht erfasst (Cooper 1994).

Von Cöster et al. (1994) werden jährliche Inzidenzraten von Hüftfrakturen auf 100.000 Einwohner in Deutschland berichtet, die für Frauen 291.3 und für Männer 110.2 betragen. Frauen erlitten damit ca. 2.5 mal so häufig Hüftfrakturen wie Männer. Ein exponentieller Anstieg der Häufigkeiten wurde für beide Geschlechter nach dem 70. Lebensjahr beobachtet. Das Lebenszeitrisiko einer 50 jährigen Frau, eine hüftgelenksnahe Fraktur zu erleiden, beträgt ca. 18 %, das Risiko, allgemein einen durch Osteoporose bedingten Bruch zu erleiden, ca. 40 %.

Tabelle 25: Alterskorrigierte, standardisierte jährliche Inzidenzrate der Schenkelhalsfrakturen für Frauen und Männer ≥ 35 Jahre (bezogen auf 100.000 Einwohner) in Deutschland im internationalen Vergleich (nach Pfeifer und Minne 2001)

	Frauen	Männer	Geschlechterverhältnis (wbl/ml)
Norwegen, Oslo	421	230	1.8
Deutschland	407	166	2.4
USA, Rochester	320	177	1.8
Schweden, Malmö	378	241	1.6
Italien	232	81	2.8
England, Trent	294	169	1.7
Japan, Tottori	108	54	2.0
China, Shenyang	87	100	0.9
Korea, Honam	41	48	0.9

Zum internationalen Vergleich wurden die alters- und geschlechtsbezogenen Häufigkeiten der Schenkelhalsfrakturen bei Personen über 35 Jahren berechnet und ähnlich dem Vorgehen von Riggs und Melton (1995) standardisiert auf die Bevölkerungsstruktur der USA in obenstehender Tabelle zusammengestellt.

Im internationalen Vergleich zeigt sich, dass die in Deutschland ermittelten Inzidenzen der Schenkelhalsfrakturen eher den Häufigkeiten in Schweden und Norwegen entsprechen als den angelsächsischen Ländern und deutlich höher liegen als beispielsweise in der Mittelmeerregion (Italien) oder Asien. Die jährliche Inzidenz der Schenkelhalsfrakturen wird sich in Deutschland nach neueren Schätzungen bis zum Jahr 2050 mehr als verdoppeln. Dies ist eine der höchsten Zunahmen innerhalb der Staaten der EU (Cooper 1994).

Periphere Skelettfrakturen haben in fast allen Fällen unmittelbare und erhebliche gesundheitliche Auswirkungen. In 97 % der Fälle führen sie zu ärztlichen Konsultationen und Behandlungen; 68 % der Patienten wurden allein ambulant, 10 % allein stationär und 19 % sowohl ambulant als auch stationär behandelt. Eine Publikation von Cooper et al. (1993) berichtet die Mortalität nach osteoporotischen Frakturen in mehreren Kohorten von Patienten nach Erstdiagnose der entsprechenden Frakturart. Die Kohorte der Wirbelfrakturpatienten bestand aus 79 Männern und 256 Frauen aller Altersgruppen (Cooper et al. 1992), die Hüftfrakturkohorte umfasste die Daten von 456 Einwohnern aller Altersgruppen (Melton III et al. 1987). Sie konnten für Patienten mit stattgehabten Hüftfrakturen eine relative 5-Jahresüberlebenschance im Vergleich zur Normalbevölkerung von 0,82 (95% CI 0,77 - 0,87) für vertebrale Frakturen eine relative 5-Jahresüberlebenschance von 0,81 (95% CI 0,70 - 0,92) bestimmen. Bei Hüftfrakturen ist die höchste Sterblichkeit in den ersten sechs Monaten nach Fraktur zu verzeichnen, die Mortalität nach vertebrealen Frakturen zeigt keinen ausgesprochenen Häufigkeitsspitzenwert. Die Mortalitätsrate 1 Jahr nach erlittener Schenkelhalsfraktur beträgt ca. 10 - 15%, ca. 30 % der Betroffenen werden versorgungspflichtig invalide (Gregg et al. 2000). Die Übrigen erleiden Einschränkungen ihrer Leistungsfähigkeit und Beweglichkeit. Hinsichtlich der Kosten einer Schenkelhalsfraktur errechneten Schürch et al. (1996) eine Summe von 50.700 Schweizer Franken, deutsche Autoren schätzten unter Berücksichtigung des unterschiedlichen Preisniveaus zwischen der Schweiz und Deutschland die Ausgaben auf 39.700 DM (= € 22.300) (Pfeifer und Minne 2001).

Zusammenfassend kommen Pfeifer und Minne (2001) zu der Feststellung, dass Schenkelhalsfrakturen eine deutlich höhere Inzidenz aufweisen als bisher angenommen und dass sie mit über 5 Milliarden DM (= € 2.56 Milliarden) Folgekosten eine erhebliche und steigende Belastung der Systeme der Sozialversicherung darstellen.

II C.2.1.3 Relation von Knochendichte und Frakturrisiko

Die Größenordnung des Zusammenhanges von niedriger Knochendichte und Fragilität und ihr Einfluß als Risikofaktor für Frakturen muss anhand von Ergebnissen aus prospektiven Kohortenstudien und Querschnittsuntersuchungen ermittelt werden.

Die Meta-Analyse von Marshall et al. (1996) analysiert getrennt prospektive Studien und Fall-Kontroll-Studien zur Untersuchung des Zusammenhanges von Frakturrisiko und Knochendichte und berechnet relative Risiken für Frakturen in Abhängigkeit von der Knochendichte. Die Verfasser erkennen eine klare Beziehung zwischen niedriger Knochendichte und erhöhter Frakturnrate, wobei die Beziehung sich exponentiell verhält, d. h.: Liegt die gemessene Knochendichte um 1 % unter dem Mittel, erhöht sich das Risiko, eine Fraktur zu erleiden, um 5 %; liegt sie um 10 % unter dem Mittel, erhöht sich das Frakturrisiko um fast 90 %.

Während Messungen an Hüfte und Wirbelsäule bessere Vorhersagewerte für Frakturen an genau diesem Messort andeuten, konnte dies für Messungen und Frakturen am Unterarm nicht dargestellt werden.

Die Autoren der Meta-Analyse kommen zu den folgenden zwei relevanten Schlussfolgerungen:

1. Die Ergebnisse demonstrieren einen Zusammenhang von niedriger Knochendichte und erhöhter Frakturnrate. Die Bestimmung eines „Schwellenwertes“ ist jedoch nicht möglich, d. h. für individuelle Personen kann keine Aussage gemacht werden, ob sie eine Fraktur erleiden werden oder nicht. Insbesondere gilt dies für die Vorhersage von Frakturen im höheren Alter aus einer Messung im mittleren Alter.
2. Aus den Ergebnissen der Metaanalyse berechnen die Autoren eine anzunehmende Reduktion der Frakturnraten um 2,5 % bzw. 20 %, wenn es gelingt, bei Frauen mit einer Knochendichte $< -1SD$ vom altersadjustierten Mittel durch therapeutische Interventionen die Knochendichte um 1 % bzw. 10 % anzuheben. Es wird jedoch betont, dass es sich hierbei um hypothetische Werte handelt, die durch prospektive Studien zu belegen sind.

II C.2.1.4 Präventions- und Behandlungsmöglichkeiten

Präventions- und Behandlungsinterventionen bei Osteoporose werden mit drei übergeordneten Intentionen eingesetzt:

- Primärprävention ist auf die Vermeidung von Krankheit ausgerichtet,
- Sekundärprävention zielt auf die Früherkennung von Störungen in einem gut behandelbaren Stadium und
- Tertiärprävention (mit fließendem Übergang zur Therapie) beschreibt Strategien, die bei bekanntem Krankheitsstatus Rezidive bzw. Verschlechterungen verhindern sollen.

Behandlungsansätze basieren auf der Annahme, dass die Knochendichte entscheidenden Einfluss auf die Frakturrate hat und dass eine Erhöhung der BMD bzw. die Verhinderung ihrer Erniedrigung die Rate an osteoporotischen Frakturen senken kann. Die möglichen Therapieformen werden hier nicht weiter beschrieben, da sie nicht Gegenstand dieses Berichtes sein sollen. Weitere Ausführungen zu medikamentösen Interventionen finden sich in der einschlägigen Literatur und im HTA-Bericht Nr. 13 „Die Rolle der Osteodensitometrie im Rahmen der Primär-, Sekundär- und Tertiärprävention/Therapie der Osteoporose“ von Lühmann et al. (2000).

II C.2.2 Hintergrund/Beschreibung der Technologie

II C.2.2.1 Einführung

Ultraschallverfahren im Bereich der Osteoporosediagnostik unterscheiden sich technologisch grundlegend von den im klinischen Einsatz verbreiteten sonografischen Verfahren.

QUS-Verfahren sind parametrische, in der Regel nichtbildgebende Untersuchungsverfahren, bei denen die Wechselwirkung von Ultraschall und Knochen analysiert wird. „Parametrisch“ bedeutet in diesem Zusammenhang, dass nicht eine subjektiv zu interpretierende (Bild-)Information, sondern ein quantitatives Messergebnis gewonnen wird, das Aufschluss über die Knocheneigenschaften geben soll. Auch die bildgebenden QUS-Verfahren (Laugier et al. 1994; Roux et al. 1996) verwenden das Bild nicht als diagnostische Informationsquelle, sondern lediglich als Hilfsmittel zur anatomisch korrekten Platzierung der Messregion.

Nach Barkmann et al. (1999) ist der Einsatz dieser Techniken zur **quantitativen** Bestimmung akustischer Parameter menschlicher Knochen relativ neu. Das Interesse an quantitativen Ultraschallverfahren zur Messung von Knocheneigenschaften entstand zum einen aus Beschränkungen konventioneller radiologischer Knochendichtemessungen als auch aus Kosten-/Nutzen-Erwägungen. Zwar beeinflusst die messbare Knochendichte zu einem erheblichen Anteil die Frakturanfälligkeit des Skeletts, aber nicht nur die Masse des Knochens, sondern auch die Zusammensetzung der Knochensubstanz (Qualität) sowie ihre Verteilung (Struktur) sind von Bedeutung. Ultraschallwellen werden nicht nur von der Materialmasse, sondern auch von anderen Eigenschaften wie Elastizität und Struktur beeinflusst und sind daher potentiell in der Lage, zusätzliche Informationen über die Knochenfestigkeit zu liefern. Außerdem sind Ultraschallgeräte kostengünstiger herzustellen und völlig frei von ionisierender Strahlung.

(Quantitativer) Ultraschall (QUS) ist in der Lage, eine Vielzahl von Informationen über die untersuchte Substanz zu liefern. Ultraschallwellen sind mechanische Wellen, die (im Gegensatz zu Röntgenstrahlen) ein Medium zur Ausbreitung benötigen. Dementsprechend bestimmen die Materialeigenschaften dieses Mediums sehr stark die Ausbreitung des Schalls. Zwei Parameter werden hauptsächlich zur Beschreibung

der Schallausbreitung im Knochen herangezogen: Die Schallgeschwindigkeit, zu meist als „Speed of Sound“ (SOS) bezeichnet und in Einheiten von m/s angegeben und die Schallabschwächung, die sog. „Breitband-Ultraschall-Abschwächung“ (engl. Broadband Ultrasound Attenuation, BUA) (Miller et al. 1993; Langton et al. 1984). Die Schallgeschwindigkeit in einem Festkörper hängt primär von Dichte und Elastizität des Stoffes ab (Shutilov 1984), während die Schwächung überwiegend bestimmt wird durch Absorption im Medium und Streuung an Inhomogenitäten (Njeh et al. 1999).

Leider sind diese prinzipiell genau bekannten Zusammenhänge zwischen Ultraschallparametern und Stoffparametern in einem inhomogenen Körper wie dem Knochen kompliziert und theoretisch nur näherungsweise berechenbar. Das gilt nicht nur für die Spongiosa mit ihrer markanten schwammartigen Struktur, sondern auch für die Kompakta, welche mechanisch und akustisch auch anisotrop ist (Katz et al. 1987), eine physikalische Eigenschaft, die sich z. B. in der hohen Biegesteifigkeit langer Knochen, aber geringer Resistenz gegen Torsion äußert.

Im Folgenden werden die unterschiedlichen Messprinzipien kommerzieller Geräte, die in Deutschland angewendet werden, erläutert und die Messparameter beschrieben.

II C.2.2.2 Messmethoden

Da die Ultraschallausbreitung im Knochen durch verschiedene Parameter beschrieben werden kann, sind eine Reihe von unterschiedlichsten Messmethoden zu verstehen. Nach Glüer et al. (1998) sind zur Zeit folgende drei Verfahren bis in die klinische Anwendbarkeit gelangt:

- Quertransmission der Spongiosa
- Quertransmission der Kortikalis
- Längstransmission der Kortikalis

Diese Verfahren sind charakterisiert durch verschiedene Ausbreitungsmodi mit unterschiedlicher Abhängigkeit von Knochenparametern.

Quertransmission der Spongiosa

Dieses Verfahren arbeitet nach folgendem Prinzip: Durchschallung eines überwiegend trabekulären Knochens (zur Zeit nur des Kalkaneus) mit Hilfe zweier sich gegenüberstehender Schallwandler, einem Sender und einem Empfänger. Unterschiede zwischen den Verfahren bestehen in der Art der Schalleinkopplung (Wasserbad, Gel) und in der Verwendung feststehender oder scannender Schallwandler.

Im folgenden ist eine Auswahl von in Europa verbreiteten Geräten aufgeführt:

Verwendung eines Wasserbades zur Schalleinkopplung:

- Achilles+, Lunar, fixierte Schallwandler
- UBIS 5000, Diagnostic Medical Systems (DMS), scannend, bildgebend
- DTU-one, Osteometer, scannend, bildgebend

Schalleinkopplung mit Ultraschallgel:

- Cuba Clinical, McCue, fixierte Schallwandler
- Sahara, Hologic, fixierte Schallwandler
- QUS-2, Metra, scannend

Bedingt durch die starke Dämpfung hoher Frequenzen im Knochen läßt sich der üblicherweise in der Sonografie benutzte Frequenzbereich über 2 MHz nicht anwenden. Als sinnvoll hat sich die Applikation eines Breitbandpulses mit einer Mittenfrequenz von 0,5 MHz erwiesen. Die Laufzeit dieses Schallimpulses sowie seine frequenzabhängige Schwächung nach Durchlaufen des Kalkaneus werden zur Berechnung zweier Parameter, der Schallgeschwindigkeit (SOS) sowie der Breitband-Ultraschallabschwächung (BUA) benutzt.

- a) SOS: Die Geschwindigkeit der Schallwelle im Knochen wird berechnet aus dem Verhältnis von Wandlerabstand und Schalllaufzeit zwischen den Wandlern. Fehlerquellen sind das unbekanntes Verhältnis zwischen Knochen- und Weichteildicke sowie variable Weichteilgeschwindigkeiten. Bei vielen Kalkaneus-Messgeräten wird diese Geschwindigkeit als Speed of Sound (SOS) angegeben (Njeh et al. 1997).
- b) BUA: Die Schwächung des Schallstrahls hängt in starkem Maße von der Frequenz ab. Im üblichen Frequenzbereich von 0,2 – 0,8 MHz steigt die Schallschwächung am Kalkaneus in erster Näherung linear mit der Frequenz an. BUA ist ein Maß für den linearen Anstieg der Schallschwächung mit der Frequenz (Einheit: dB/MHz) (Langton et al. 1984).

Studien an Proben von trabekulärem Knochen zeigen eine Abhängigkeit von SOS und BUA sowohl von der Dichte als auch von Strukturparametern (Glüer et al. 1994). Im klinisch relevanten Bereich (medio-laterale Durchschallung des Kalkaneus) allerdings erklärt die Knochendichte bei identischen Messregionen ca. 90 % der QUS-Parameter, womit hier wenig Raum für Struktureinfluss bleibt (Hans et al. 1999). Es ist noch Gegenstand der Forschung, ob unterschiedliche Änderungen von Dichte und Struktur, bedingt durch verschiedene Erkrankungsarten oder während einer Therapie, mit Ultraschall feiner als mit einer reinen Knochendichtemessung erfasst werden können (Barkmann et al. 1999).

Quertransmission der Kortikalis

Dieses Verfahren arbeitet ebenfalls mit zwei gegenüberliegenden Schallwandlern und Querdurchschallung des Knochens. Als einzige Knochen werden klinisch zur Zeit die proximalen Phalangen gemessen, und zwar an der distalen Metaphyse (*DBM Sonic Bone Profiler, IGEA*). In diesem Bereich befindet sich kortikale als auch trabekuläre Knochensubstanz. Die Geschwindigkeit der Ultraschallimpulse einer Frequenz von 1,25 MHz im Knochen wird auch hier aus dem Verhältnis zwischen Wandlerabstand und Schalllaufzeit berechnet.

Eine Besonderheit ist die Berechnung einer amplitudenabhängigen Schallgeschwindigkeit (Amplitude Dependent Speed of Sound, **AD-SoS**) als Standardparameter, welcher Schallgeschwindigkeit und –schwächung in einem Parameter miteinander kombiniert. Ein weiterer Parameter beschreibt die Form des empfangenen Signals (Ultrasound Bone Profile Index, **UBPI**).

Die Geschwindigkeit der Schallwelle wird in starkem Maße durch die Fingermorphometrie beeinflusst. Sowohl endostale Resorption mit kortikalem Massenabbau, als auch eine Zunahme der kortikalen Porosität erniedrigen die Schallgeschwindigkeit. Die mit endostaler Resorption verbundene Vergrößerung des endomedullären Kanals vermindert aber vor allem die Amplitude des empfangenen Signals.

Die amplitudenabhängige Schallgeschwindigkeit (**AD-SoS**), welche von Amplituden- und Geschwindigkeitsänderungen beeinflusst wird, reagiert daher sensitiv auf all diese mit Alterung und Krankheit einhergehenden Änderungen. Diese Messgröße wird am nicht-gewichttragenden Knochen der Fingerphalangen (Alenfeld et al. 1998), der am Messort aus vorwiegend kortikalem Knochen besteht, bestimmt. Allerdings ist der diagnostische Wert dieses Messparameters noch nicht abschließend geklärt. (Glüer 1999).

Längstransmission der Kortikalis

Bei der Längstransmission wird die Schallausbreitung in einer dünnen Schicht unter der Knochenoberfläche gemessen. Zur Zeit gebräuchliche Messorte sind Tibia, Phalanx, Metatarsale und Radius (*OmnisenseTM, Sunlight Ultrasound Technologies*). Während die Quertransmission nur bei Knochen eingesetzt werden kann, die von zwei gegenüberliegenden Seiten zugänglich sind, läßt sich die Längstransmission prinzipiell an allen Stellen des Skeletts bestimmen, welche von einer Seite ohne zu dicke Weichteilschichten erreichbar sind. Im klinischen Gerät beherbergt ein Gehäuse Sender und Empfänger sowie weitere Schallwandler, mit denen eine Kompensation des Weichteileinflusses durchgeführt wird.

Schallwellen, die über weiches Gewebe in den Knochen eingekoppelt werden, treten unter verschiedenen Brechungswinkeln in die feste Materie ein. Abhängig vom Einfallswinkel breiten sich die Wellenanteile im Knochen dann unter verschiedenen Winkeln weiter aus. Die Schallanteile, die unter einem bestimmten Winkel („kritischer

Winkel“) auf den Knochen treffen, bewegen sich innerhalb des Knochens parallel zu seiner Oberfläche fort. Dabei wird laufend Schallenergie in das weiche Gewebe zurückgesandt, welche dann vom Empfänger detektiert werden kann. Mit Hilfe einer geschickten Kompensation des Weichteileinflusses (Barkmann et al. 1999) lässt sich dann aus der bekannten Strahlgeometrie die Schallgeschwindigkeit im Knochen berechnen. So lange die Dicke der Kortex einige Millimeter überschreitet, hängt die Schallgeschwindigkeit nicht von der Kortexdicke ab, sondern gibt überwiegend die Elastizität des Knochens wieder. Zunehmende Porosität vermindert ebenfalls ihren Wert.

Die in Längstransmission bestimmte SOS an nicht gewichttragenden Knochen, wie den Fingerphalangen und dem Radius, reflektieren ausschließlich kortikale Eigenschaften des Knochens (Hans et al. 1999). Dieses Verfahren wird auch an der Tibia eingesetzt, jedoch sind die Ergebnisse so wenig überzeugend, dass dieser Anwendung gegenwärtig keine klinische Bedeutung zukommt.

Nach Glüer et al. (1999) sind die Differenzen in den Verfahren bezüglich des Messortes, der untersuchten Knochenart und der Art der Signaltransmission von großer Bedeutung. Bei den gegenwärtig kommerziell erhältlichen Geräten sind insgesamt 5 Ultraschallmessgrößen (4 Schallgeschwindigkeiten SOS und einen BUA-Parameter) zu unterscheiden, die in nachstehender Tabelle 26 aufgeführt sind.

Über die Grundparameter hinaus werden auch weitere Kombinationsparameter errechnet. Am verbreitetsten ist nach Glüer der Knochensteifigkeitswert („Stiffness“), ein aus BUA und SOS zusammengesetzter gewichteter Mittelwert. Andere Kombinationsparameter, die ähnlich definiert sind, heißen Quantitativer Ultraschallindex (QUI), „soundness“ usw. Sie alle liefern keine neuen, über die Grundparameter SOS und BUA hinausgehende Informationen (Glüer et al. 1999).

Die Kombinationsparameter sind praktisch, weil sie die Werte zusammenfassen und sich grundsätzlich besser zur Verlaufskontrolle (z. B. aufgrund der geringeren Temperaturempfindlichkeit (Morris et al. 1997)) eignen. Für die Frakturrisikoabschätzung bieten sie nach allen bekannten Ergebnissen keine nennenswerten Vor- oder Nachteile. Als gewichteter Mittelwert sollen sie etwas schlechter als der jeweils bessere der Einzelparameter geeignet sein, auf der anderen Seite sind sie wiederum etwas stabiler.

In der Zusammenfassung erläutert Glüer (1999) in seiner Übersichtsarbeit über Quantitativen Ultraschall, dass für die meisten Anwendungen die Unterscheidung der vermessenen Knochenart (gewichtstragend oder nicht, spongiös oder kortikal) ebenso wichtig wie die Wahl des Parameters und des Messverfahrens ist. Gerade unter medikamentöser Therapie und bei sekundären Osteoporosen können erhebliche Unterschiede auftreten.

Tabelle 26: Messcharakteristika der quantitativen Ultraschallverfahren

Parameter	Messort	Gewichts- tragend	Kortikalis- einfluss	Spongiosa- einfluss	Transmis- sion	Amplitu- den- abhängig
BUA	Kalkaneus	Ja	Minimal	Ja	Quer	Nein
SOS	Kalkaneus	Ja	Minimal	Ja	Quer	Nein
ADSoS	Phalangen	Nein	Ja	Variabel	Quer	Ja
SOS	Phalangen, Radius	Nein	Ja	Nein	Längs	Nein
SOS	Tibia	Ja	Ja	Nein	Längs	

II C.2.2.3 Genauigkeit, Richtigkeit und Präzision

Ein zentraler Begriff in der medizinischen Diagnostik ist der sogenannte „Test“. Dieser wird in der Medizin unterschiedlich verwendet. Zum einen wird er benutzt, um die Messung (den Messvorgang) eines Merkmals zu bezeichnen, zum anderen wird das Testergebnis als Hinweis (Indikator) für eine Erkrankung gesehen.

Die Bewertung eines Messverfahrens allein reicht im Allgemeinen nicht für die Beurteilung des diagnostischen Tests aus. Ein schlechtes Messverfahren für einen guten diagnostischen Test (Indikator) kann durchaus wertvoller zur Erkennung einer Erkrankung sein als ein gutes Messverfahren für einen schlechten diagnostischen Test.

Nach Lange et al. (1994) ist die Kenntnis der Zuverlässigkeit der zur Diagnostik benutzten Methoden eine notwendige aber nicht hinreichende Voraussetzung für den weiteren Einsatz als diagnostischer Test. Der Begriff der Zuverlässigkeit umfasst verschiedene Komponenten, die im Nachfolgenden definiert und näher erläutert werden sollen. Arbeiten, die sich mit der Bestimmung dieser „Zuverlässigkeitskriterien“ einer Messmethode befassen, werden als Phase-I-Studien bezeichnet.

Bei Köbberling et al. (1991) wird der Begriff Genauigkeit als Oberbegriff für die Übereinstimmung von Messwert und "wahrem Wert" bezeichnet.

Ein Verfahren arbeitet umso genauer, je weniger der Mittelwert einer Messserie vom wahren Wert abweicht (Richtigkeit) und je enger sich die Einzelwerte um den Mittelwert scharen (Präzision).

In dem vorliegenden Bericht zur Frakturrisikoabschätzung mit Hilfe der quantitativen Ultraschallverfahren (QUS) müssen zunächst die folgenden Fragen beantwortet werden:

- Die Frage nach der Richtigkeit und Präzision: Wie korrekt und präzise lässt sich ein Parameter messen?

- Die Frage nach der diagnostischen Wertigkeit (prognostische Validität): Wie wahrscheinlich ist es, dass bei einer bestimmten Ausprägung des Parameters eine Krankheit/Frakturgefährdung vorliegt?

Die **Präzision (Reproduzierbarkeit)** ist ein Maß dafür, wie stark Messergebnisse aus wiederholten Messungen unter gleichen Bedingungen streuen. Das Ausmaß der „Streuung“ wird beschrieben durch den Variationskoeffizienten CV, der sich berechnet als Quotient aus Standardabweichung und Mittelwert der Abweichungen in Prozent.

Nach Lange et al. (1994) ergibt sich die praxisrelevante Reproduzierbarkeit eines Verfahrens erst, wenn das Ausmaß der Annäherung zwischen Ergebnissen von Messungen derselben Messgröße unter folgenden zum Teil unterschiedlichen Bedingungen ermittelt wurde:

- Kompatible Messverfahren (differente Methoden)
- Anwendungsbedingungen (andere Umweltbedingungen)
- Beobachter (Untersucher)
- Messorte (andere Arbeitsplätze)
- Messeinrichtungen (verschiedene Geräte)
- Zeiten (Abstände zwischen den Messungen)

Eine Änderung ist statistisch signifikant messbar, wenn sie das 2.8-fache des Präzisionsfehlers übersteigt (95 % Konfidenzintervall, zweiseitiger Test) (Blake & Fogelman 1997; Nguyen et al. 1993).

Verlaufskontrollen in zu kurzen Zeitabständen (d. h. in der Regel unter einem Jahr mit Ausnahme von speziellen sekundären Osteoporosen) führen leicht zu Fehlinterpretationen, da minimale Änderungen (von unter 2.8 x Präzisionsfehler) rein statistischer Natur sein können und keinen Aufschluss über tatsächliche Knochendichteveränderungen gestatten (Glüer 2000; Blake & Fogelmann 1997; Nguyen et al. 1993): Die Wahrscheinlichkeit, dass eine beobachtete Änderung nicht stattgefunden hat, beträgt in diesem Fall 5 %. Um einen Trend zu erkennen muss die erwartete Änderung nur noch das 1.8-fache des Präzisionsfehlers übersteigen (10 % Irrtumswahrscheinlichkeit, einseitiger Test).

Dies hat also direkte Auswirkungen auf einen sinnvollen Abstand zwischen zwei Messungen, um z. B. den Therapieerfolg hinsichtlich einer Veränderung der Knochendichte zu kontrollieren.

In der folgenden Tabelle werden die unterschiedlichen Verfahren und Geräte zur Knochenmessung mit Hilfe des Quantitativen Ultraschalls aufgelistet und die Präzision angegeben.

Tabelle 27: QUS-Verfahren und Geräte, Präzision nach Hans (1997)

Ultraschall-system	Kopplungs-medium	Parameter	Präzision (CV)	Bemerkungen
Kalkaneus einzeln, fixed point QUS Systeme				
Lunar Achilles Plus	Wasserbad	BUA SOS Stiffness	0.8 - 2,5 % 0.2 - 0.4 % 1.0 - 2.0 %	Time of flight velocity Heel width ‚constant‘
McCue CUBA Clinical	Gel	BUA SOS	1.5 - 4.0 % 0.2 - 0.6 %	Heel velocity (Weichteilgewebe und Knochen)
Hologic Sahara	Gel	BUA SOS QUI	0.8 - 2.5 % 0.2 - 0.4 % 1.0 - 2.0 %	Heel velocity (Weichteilgewebe und Knochen)
Kalkaneus bildgebende QUS Systeme				
DMS UBIS 3000	Wasserbad	BUA SOS	0.8 - 2.5 % 0.2 - 0.4 %	Scannend, bildgebendes System
Osteometer DTU-1	Wasserbad	BUA SOS	0.8 - 2.5 % 0.2 - 0.4 %	Scannend, bildgebendes System
Tibiale QUS Systeme (semi-reflection mode)				
Myriad Soundscan 2000	Gel	SOS	0.2 - 1.0 %	Geschwindigkeit entlang der Kortikalisoberfläche
Phalangeale QUS Systeme				
IGEA DBM Sonic 1200	Gel	AD-SoS	0.5 - 1.0 %	Amplitudenabhängige Geschwindigkeit
Multi-site QUS Systeme (semi-reflection mode); (Kalkaneus, Tibia, Phalangen, Radius, Ulna, Trochanter etc.)				
Sunlight Omnisense one	Gel	SOS	0.2 - 1.0 %	Geschwindigkeit entlang der Kortikalisoberfläche
Nach: Hans D. How can we measure bone quality? Bailliere's clinical rheumatology: 11 (3):495-515; 1997				

CV: Variationskoeffizient, BUA: Breitband-Ultraschall-Abschwächung, QUI: Quantitativer Ultraschall-Index, SOS: Schallgeschwindigkeit, AD-SoS: Amplituden-abhängige Schallgeschwindigkeit.

Die Präzision der QUS-Verfahren variiert abhängig von den gemessenen Parametern, des Messortes und des Systems. Der CV % von BUA variiert von 0.8 bis 4.0 %, der CV % von SOS von 0.2 bis 1.0 %. Der Variationskoeffizient CV % von SOS ist somit ca. viermal kleiner als der von BUA.

Die Langzeit-Präzision der QUS-Verfahren wird in der Literatur als annäherungsweise zwei- bis dreimal so groß wie die für die DXA-Techniken angegebenen Werte beschrieben (Greenspan et al. 1997; Van Daele et al. 1998; Sahota et al. 2000).

Standardisierter Präzisionsfehler: Zur besseren Beschreibung der klinischen Wertigkeit des Messverfahrens muss zusätzlich zur Reproduzierbarkeit (beschrieben durch den Präzisionsfehler) die Ansprechrate (Änderungsrate des Wertes im zeitlichen Verlauf) des Verfahrens in Rechnung gestellt werden. Der Präzisionsfehler alleine ermöglicht keine Beurteilung der klinischen Signifikanz. Eine Messmethode mit großem Präzisionsfehler an einem Messort mit hoher Ansprechrate kann die gleiche klinische Signifikanz besitzen wie eine Methode mit geringem Präzisionsfehler an einem Ort mit geringer Ansprechrate.

Typische Ansprechraten eines „normalen“ Knochenmasseverlustes (gemessen in den Standardverfahren DXA und QCT) postmenopausaler Frauen sind pro Jahr:

DXA der Wirbelsäule oder des Kalkaneus	1 – 2 %
DXA des proximalen Femurs	0,5 - 1 %
QCT der Wirbelsäule	2 - 3 %

Für Verlaufskontrollen ist also eine möglichst hohe Reproduzierbarkeit erforderlich, um geringe Veränderungen zuverlässig bestimmen zu können.

In neueren Studien (Iki et al. 1999, He et al. 2000, Njeh et al. 2000; Frost et al. 1999; Frost et al. 2000; Sahoto et al. 2000) wird zusätzlich der standardisierte Präzisionsfehler kalkuliert, um eine bessere Vergleichsgröße zur Wertigkeit einzelner Osteometrieverfahren zu erhalten. Die isolierte Angabe von Richtigkeits- oder Präzisionsfehlern, auch wenn sie in Prozent ausgedrückt sind, ist ohne Angabe der korrespondierenden Ansprechrate irreführend. Ein Verfahren kann kleine prozentuale Fehler aufweisen und dennoch untauglich sein, dann nämlich, wenn die Ansprechrate noch kleiner ist. Gerade bei Ultraschallverfahren ist dies von besonderer Bedeutung, da die Prozentwerte von Schallgeschwindigkeitsmessungen und Absorptionsmessungen (SOS und BUA) absolut unvergleichbar sind. Ein Verfahren für SOS-Werte muss z. B. eine um den Faktor 5 - 10 bessere Präzision in Prozent aufweisen, um auf Grund der kleineren Ansprechrate eine ähnliche longitudinale Ansprechempfindlichkeit (=Sensitivität einer Messmethodik zur Verlaufskontrolle) zu erzielen. Als bessere Vergleichsgröße beschreiben Glüer et al. (1995) deshalb den standardisierten Präzisionsfehler wie folgt:

Standardisierter Präzisionsfehler sPE % = Präzisionsfehler CV % × Ansprechrate von DXA der LWS / Ansprechrate des Verfahrens

Da der standardisierte Präzisionsfehler von DXA der Lendenwirbelsäule als einem anerkanntermaßen empfindlichen Verfahren bei etwa 1 - 1,5 % liegt, könnte ein zu vergleichendes Verfahren, das einen standardisierten Präzisionsfehler in gleicher Größenordnung aufweist, als ähnlich empfindlich und zur Verlaufskontrolle geeignet eingestuft werden.

Frost et al. (1999) verglichen Kurz- und Langzeitpräzisionen einiger QUS-Verfahren und fanden heraus, dass die gemessenen Werte über einen Zeitraum von 10 Mona-

ten doppelt so hoch waren wie die Werte für die Kurzzeitpräzision. Dies bestätigt ältere Angaben von Roux et al. (1996). Die neueste Studie von Frost, Blake und Fogelman (2000) zeigt zwar bei der standardisierten Kurzzeitpräzision bei bildgebenden QUS-Verfahren am Kalkaneus eine deutliche Verbesserung, die Langzeitpräzision weist jedoch keine Änderung auf, sodass die Autoren zunächst weitere Forschung in der Entwicklung der QUS-Verfahren für erforderlich halten.

Man kann bei der Zuverlässigkeit einer Messung eine zweite Komponente unterscheiden, die sogenannte „Richtigkeit“.

Die **Richtigkeit (engl. accuracy)** ist ein Maß dafür, wie nahe das Messergebnis an den wahren Wert herankommt. Sie setzt Klarheit über die zu messende biologische Größe und deren Beziehung zum eigentlichen Zielparameter, also über die Validität des Messverfahrens voraus. Eine Messung ist nach allgemeiner Definition valide, wenn sie das misst, was sie messen soll. Am Beispiel der Knochendichtemessung mit DXA: Zielparameter ist die in vivo nicht zu bestimmende Bruchfestigkeit des Knochens, die zu einem wesentlichen Teil durch die so genannte Knochendichte bestimmt wird. Diese wird schließlich als Knochenmineralsalzgehalt pro Flächen- oder Raumeinheit physikalisch gemessen. Die Messung ist abhängig auch von systematischen Fehlern, d. h. unter anderem von der Kalibrierung (Eichung) des Systems. Das bedeutet: Bei gleicher Technik am gleichen Messort kann die Vergleichbarkeit der Knochendichtewerte verschiedener Scanner und Hersteller nur durch entsprechende Standardisierungsverfahren sichergestellt werden. Systematische Fehler werden durch Vergleich mit Normalwerten und entsprechenden Korrekturen teilweise ausgeglichen.

Studien zur Richtigkeit werden entweder am Phantom („in vitro“) oder an Leichenknochen („in vivo“) mit anschließender Veraschung vorgenommen. Die Richtigkeit wird dann zumeist als Standardfehler einer linearen Regressionsgeraden oder als Anteil „nicht erklärter Varianz“ ($1 - r^2$) angegeben (jeweils in %) (WHO 1994); als Synonym wird häufig der Begriff „Genauigkeitsfehler“ bzw. „Richtigkeitsfehler“ verwendet.

Angaben zur Validität und Richtigkeit der QUS-Messverfahren finden sich in der von uns bis einschließlich Juni 2001 durchgeführten Literaturrecherche nicht; ein Grund mag sein, dass sich die komplexe Ausbreitung von Schallwellen im inhomogenen Knochen bisher weitgehend der Vorstellung entzieht. Dies deckt sich auch mit den Angaben von Homik und Hailey im HTA-Bericht aus dem Jahre 1998.

Eine Reihe von Studien wurde durchgeführt mit der zentralen Fragestellung: Welche biologische Eigenschaft oder Größe des Knochens wird exakt mit dem QUS-Verfahren gemessen? Die Regression zwischen BMD-Messwerten und BUA-Parametern ist für trabekulären Knochen anders als für kortikalen Knochen (Palmer & Langton 1987). BUA wird hauptsächlich am Kalkaneus gemessen, der aus vorwiegend trabekulärem Knochen besteht. Messungen von QUS und BMD am trabekulären Knochen weisen einen hohen Korrelationskoeffizienten mit R-Werten über 0.8 auf (Laugier et al. 1997).

Die Korrelation von QUS-Parametern mit densitometrischen Messungen wurde vielfach durchgeführt. Die R-Werte dieser Untersuchungen sind abhängig vom Alter und Geschlecht und somit heterogen. Nachstehende Tabelle – modifiziert nach der Literaturübersicht von Prins et al. (1997) - zeigt zusammenfassend die gewichteten Mittelwerte sowie die Wertebereiche für drei verschiedene Messorte.

Tabelle 28: Korrelationskoeffizienten von BMD und QUS (nach Prins et al. 1997)

	BMD Schenkelhals	BMD Wirbelsäule	BMD Kalkaneus
BUA Kalkaneus	0.47 (0.30-0.87)	0.47 (0.32-0.83)	0.66 (0.53-0.74)
SOS Kalkaneus	0.52 (0.35-0.73)	0.53 (0.33-0.77)	0.47 (0.44-0.50)
SI Kalkaneus	0.61 (0.45-0.77)	0.62 (0.43-0.80)	

Die Werte basieren auf alten und neueren Studien (1990 - 1996), wobei sich in den neueren Veröffentlichungen die Tendenz zu höheren Korrelationen abzeichnet.

Die (in der obigen Tabelle nicht enthaltenen) Ergebnisse einer Studie von Chappard et al. (1997) berichten einen Korrelationskoeffizienten in der Kalkaneusregion von 0.90.

Der Korrelationskoeffizient für SOS-Parameter wird in der Studie von Kang und Speller (1998) bei Verwendung des Lunar Achilles Systems mit 0.82 angegeben.

Vor diesem Hintergrund könnte man zu der Schlussfolgerung kommen, dass der quantitative Ultraschall lediglich einen Teil der BMD-Messungen reflektiert. Die Knocheneigenschaften beinhalten zum einen Materialeigenschaften der Knochenmatrix und die Mikroarchitektur des trabekulären Knochens (Einhorn 1992). Die Untersuchungen und Forschungen in den letzten Jahren haben gezeigt, dass QUS-Parameter streng mit BMD und mit trabekulären Strukturen korrelieren. Es bleibt unklar, ob QUS irgendwelche zusätzlichen Informationen zur Knochenqualität liefern kann.

Es gibt viele potentielle Fehlerquellen bei in-vivo-Ultraschallmessungen am Knochen, z. B. dass den Knochen umgebende weiche Gewebe und die Positionierung des Knochens relativ zu den Schallwandlern.

Richtigkeit und Präzision werden beeinflusst durch Weichteileigenschaften wie die Gewebstemperatur und die Einkopplung des Schalls in die Haut (Njeh et al. 1999).

Einige Kalkaneus-Geräte verwenden zur Schallankopplung ein Wasserbad, in dem sowohl das Fersenbein als auch die Schallwandler eingetaucht sind. Andere Geräte benutzen Gel als Kopplungsmedium. Ein weiterer Unterschied zwischen den Geräten besteht darin, dass feststehende Wandler benutzt werden oder durch Abscannen ein Bild des Kalkaneus erzeugt werden kann. Während Geräte mit feststehenden Wandlern je nach Knochengröße und Weichteildicke unterschiedliche anatomische Regionen messen, bieten bildgebende Geräte den Vorteil, dass anatomisch entsprechen-

de Regionen auch in Füßen unterschiedlicher Form und Größe reproduzierbar gewählt werden können.

Barkmann et al. (1999) führen weiter aus, dass die enge Kopplung des Ultraschalls an das durchschallte Medium auch einige Einschränkungen in der Anwendbarkeit der Methode bewirkt. So wird Ultraschall an Grenzflächen zwischen Medien mit unterschiedlichen Impedanzen reflektiert, eine Eigenschaft, welche in der Sonografie zur Bilderstellung genutzt wird. Fast vollständige Reflexion und damit praktisch verschwindende Durchschallung tritt dann auf, wenn diese Impedanzunterschiede sehr groß sind. Dies ist immer dann der Fall, wenn Gase im Körper vorhanden sind. Da auch Weichteilgewebe die Ausbreitung beeinflusst, ist eine quantitative Messung dann schwierig, wenn ein ungünstiges Verhältnis zwischen Knochendicke und Weichteilgewebe vorliegt. Aus diesen Gründen ist bisher noch kein Verfahren bekannt geworden, mit dem Wirbelkörper und Oberschenkelhals akustisch vermessen werden könnten. Auch in den Messungen am Kalkaneus machen sich Eigenschaften der Weichteilschichten, wie z. B. ihre Temperatur, als Fehlerquelle bemerkbar.

Dies mag ein Grund dafür sein, dass die Präzision noch unbefriedigend ist, und daher Ergebnisse von Verlaufkontrollmessungen vorsichtig zu interpretieren sind (Barkmann et al. 1999).

Als Schlussfolgerung aus dem Vorhergesagten empfehlen Iki et al. (1999) für alle Anwender unterschiedlicher QUS-Systeme:

- Überwachung und –prüfung der gerätespezifischen Präzision
- Gebrauch des Durchschnittswertes zweier Messungen
- Benutzung des gleichen Gerätetypes bei der Verlaufskontrolle
- Konstante Messort-Temperaturen durch konstante Raumtemperaturen im ganzen Jahr

Prognostische Validität

Bisher wurden Fragen der Kriteriums- und konkurrierenden Validität von quantitativen Ultraschallverfahren (QUS) diskutiert. Eine entscheidende Frage im Kontext einer Sekundär- oder Tertiärprävention ist aber, ob ein Messverfahren in der Lage ist, das Risiko für ein zukünftiges unerwünschtes Ereignis zu quantifizieren. Damit ist die Frage nach der prognostischen Validität von QUS gestellt. QUS wäre prognostisch valide, wenn es mit seiner Hilfe gelänge, das Eintreten von Frakturen sicher vorherzusagen oder wenigstens Personen mit einem überdurchschnittlichen Frakturrisiko zu identifizieren, um ihnen ein effektives präventives Mittel anbieten zu können. Zur Quantifizierung von Risiken eignen sich u. a. absolute, relative und attributable Risiken sowie Odds Ratios.

II C.2.2.4 Technologiegebrauch in Deutschland

Das hier dargestellte Verfahren der quantitativen Ultraschalluntersuchung (QUS) gewann in den letzten Jahren als alternative Technik zu den konventionellen Densitometrieverfahren zunehmend an Bedeutung. Da das Verfahren immer noch als in Entwicklung befindlich anzusehen ist, erfordert sein Einsatz eine gründliche Ausbildung der Untersucher, umfassende Qualitätssicherungsmaßnahmen und eine strenge Indikationsstellung mit klaren Zielsetzungen. Am ehesten gegeben sind diese Voraussetzungen in osteologisch ausgerichteten radiologischen Abteilungen.

Die Markteinführung wurde vor allem durch die Gerätehersteller und weitere, im Bereich Osteoporoseversorgung produzierende Firmen initiiert und voran getrieben. Unterstützt wurde diese Entwicklung nicht unerheblich durch den Beschluss des Bundesausschuss der Ärzte und Krankenkassen, die Osteodensitometrie als Kassenleistung erst nach bereits eingetretener klinischer Fraktur im Sinne einer Tertiärprävention zu übernehmen. Hierdurch gewann die Möglichkeit, ein alternatives, „völlig schmerzfreies, schnell durchzuführendes und insbesondere strahlungsfreies“ diagnostisches Verfahren anzubieten, vor allem auch für Leistungserbringer erheblich an Attraktivität. Die Anschaffung eines QUS-Gerätes kostet den Leistungsanbieter ca. € 15.000.

Zur Zeit wird die sonografische Abschätzung des Frakturrisikos in Praxen von niedergelassenen Fachärzten (überwiegend Gynäkologen und Orthopäden) sowie zunehmend auch in Apotheken als Früherkennungsmaßnahme angeboten und durchgeführt. Die Patienten werden per Aushang, Medien oder Internet darüber informiert, dass insbesondere Frauen ab dem 40. Lebensjahr regelmäßig, d. h. einmal pro Jahr, einen Osteoporose-Check up per Ultraschall durchführen lassen sollten.

Da die Knochendichtemessung per Ultraschall ebenfalls nicht zum Leistungskatalog der gesetzlichen Krankenkassen gezählt wird, werden die Kosten für diese Untersuchung als individuelle Gesundheitsleistung (IGeL) nach der Gebührenordnung für Ärzte (GOÄ) mit ca. DM 60,-- (\approx € 30,--) in Rechnung gestellt.

Laut persönlicher Anfrage bei den einzelnen Hersteller-Firmen befinden sich z. Zt. etwa 400 DBM Sonic Bone Profiler Geräte, ca. 200 Omnisense 7000S Geräte, ca. 15 Pegasus Geräte und ca. 1.400 Lunar Achilles Geräte auf dem deutschen Markt. Das letztgenannte gilt als das wohl am häufigsten untersuchte und älteste Ultraschallgerät.

II C.3 Forschungsfragen

Vor dem Hintergrund der schnellen und zunehmenden Verbreitung dieser noch jungen Technologie in der Osteoporosediagnostik sollen deren Stellenwert und Wertigkeit in diesem HTA-Schnellverfahren untersucht werden. Dabei liegt der Schwerpunkt auf der Zusammenfassung der Ergebnisse der Frakturrisikoanalyse.

Im modifizierten Delphi-Verfahren (vergl. SOP 1) wurden folgende Fragestellungen als vordringlich zu bearbeiten identifiziert:

1. Wie valide lässt sich das Frakturrisiko mit Hilfe der quantitativen Ultraschall-Verfahren im Vergleich zum DXA-Verfahren abschätzen (Prädiktive Validität: Relatives Frakturrisiko im Vergleich zu Personen mit Normalwerten)?
2. Welches der unterschiedlichen QUS-Verfahren ist hierzu am besten geeignet?
3. Sollte die Analyse ergeben, dass die quantitative Ultraschalldiagnostik eine valide und zuverlässige Abschätzung des Frakturrisikos erlaubt: Welche qualitätssichernden Maßnahmen sind dann vor einem Einsatz in der ärztlichen Praxis zu realisieren, um zu reliablen und unverzerrten Ergebnissen zu kommen?
4. Welchen weiteren Forschungsbedarf zur Klärung des Stellenwertes der Ultraschalldiagnostik im Rahmen der Frakturprävention gibt es?
5. Wie wären die unter 1 – 3 ermittelten Ergebnisse zur Patienteninformation und zum Verbraucherschutz zu nutzen?

II C.4 Methodik des HTA-Berichtes

II C.4.1 Datenquellen und Recherchen

Die Literaturrecherche gliederte sich in zwei Abschnitte:

- Berichte von HTA-Institutionen: Nach HTA-Berichten und Übersichtsartikeln (systematische Reviews und Metaanalysen) wurde in den Datenbanken MEDLINE (MED66), Health Star (HE75), der Cochrane Library und der ISTAHC-Datenbank recherchiert.
- In einem zweiten Schritt wurde in den Jahrgängen 1998 bis 2/2001 der Datenbanken Medline, Healthstar, und Cochrane Library nach Publikationen recherchiert, die aufgrund ihres Publikationsdatums in dem gefundenen HTA-Bericht nicht berücksichtigt werden konnten.

Suchbegriffe und Strategien für alle Datenbankrecherchen sind im Anhang dokumentiert. Eine ausführliche Darstellung der Vorgehensweise im Rahmen eines Kurz-HTA-Berichtes befindet sich im Methodenmanual.

II C.4.2 Bewertung der Informationen

Die aus den elektronischen Literaturrecherchen erhaltenen Publikationen wurden manuell nach folgenden Kriterien weiter selektiert:

- Aus Titel oder Abstract der Arbeit musste hervorgehen, dass in den Leitfragen aufgeführte Aspekte in der Publikation behandelt wurden.
- Es wurden nur Publikationen in englischer, deutscher, französischer oder spanischer Sprache berücksichtigt.

Die Titel der Zitate sowie die Abstracts wurden von 2 unabhängigen Untersuchern manuell selektiert nach den folgenden Ein- und Ausschlusskriterien:

Tabelle 29: Relevanzkriterien für die Literaturselektion

Relevanz	Kriterien
Zentral	Ultraschallmessung, Frakturrisiko, prospektive Studie
Peripher	Ultraschall und DXA, Querschnittsstudie, Ultraschall, populationsbezogene Stichprobe: Normwerte, Frakturpatienten im Vergleich zu Personen ohne Frakturen
Ausschluss	Sekundäre Osteoporosen; in vitro Studien, Tierstudien, thematisch abweichend, Briefe, Editorials, Fallberichte; Sprache nicht Englisch, Deutsch, Französisch, Spanisch; eindeutig unsystematische Übersichtsartikel
Unklar	anhand der Angaben im Abstract nicht eindeutig zuzuordnen

Die Dokumentation der methodischen Qualität wurde anhand neu entwickelter Stufenpläne/Checklisten für systematische Übersichtsarbeiten und Prognosestudien vorgenommen. Diese Bewertungsinstrumente wurden einer Reliabilitätstestung unterzogen; auch hier findet sich eine ausführliche Dokumentation im Methodenmanual (s. SOP 3).

II C.5 Ergebnisse

Die Ergebnisdarstellung erfolgt für den HTA-Bericht aufgrund der Komplexität gesondert, für die Prognosestudien sowohl in inhaltlicher Zusammenfassung als auch in tabellarischer Form. Im Folgenden wird zunächst der HTA-Bericht vorgestellt, danach die systematischen Reviews zur Thematik diskutiert und abschließend relevante prospektive Kohortenstudien aus dem Zeitraum 1998 – 2001 analysiert.

II C.5.1 HTA-Bericht: Homik & Hailey (1998)

Im Rahmen der Literaturrecherche wurde ein HTA-Bericht gefunden, der den Einschlusskriterien entsprach. Die kanadische Publikation aus 1998 wird im folgenden nach Bezugsrahmen, konkreter Fragestellung, verwendeter Methodik, Ergebnissen und Schlussfolgerungen dargestellt. Abschließend erfolgt die Beurteilung der Übertragbarkeit der Ergebnisse auf Verwendung im deutschen Gesundheitssystem.

Homik J und Hailey D: Quantitative ultrasound for bone density measurement. Alberta Heritage Foundation for Medical Research, Canada, HTA 11; 1998

a) Dokumenttyp und Bezugsrahmen

Das Ziel der Alberta Heritage Foundation ist die Förderung von nationalen (Kanada/Alberta) und internationalen Forschungsvorhaben. Es werden Informationen und Daten für Entscheidungsträger im Gesundheitswesen sowie Empfehlungen zur Prävention und Therapie von Erkrankungen zur Verfügung gestellt. Die vorliegende Publikation wurde im Rahmen eines Health Technology Assessment-Programmes erstellt. (Weitere Informationen sind unter: <http://www.ahfmr.ab.ca/hta/hta-publications> abrufbar).

Bis kurz vor der Erstellung des kanadischen HTA-Berichtes im Jahre 1998 wurde das Quantitative Ultraschallverfahren lediglich im Kontext von Forschungsvorhaben eingesetzt. Die U.S. Food and Drug Administration (FDA) erteilte am 1. Juli 1998 in den USA die Genehmigung zum Gebrauch von 2 (Hologic Sahara, Lunar Achilles) der insgesamt 5 hier bewerteten Geräte zur Osteoporosediagnostik. Es wurde erwartet, dass das Interesse an dieser neuen Technologie bei Verbrauchern und Entscheidungsträgern auch im kanadischen Gesundheitswesen steigen wird. Vor diesem Hintergrund wurde die vorliegende Verfahrensbewertung durchgeführt.

b) Konkrete Fragestellung

Ziel der Bewertung war es, eine Verfahrenseinschätzung für Kanada bezüglich des quantitativen Ultraschalls (QUS) durchzuführen. Dabei wurde besonders Wert gelegt auf die Frakturrisikoabschätzung im Vergleich zum sogenannten Goldstandardverfahren (DXA-Messung) und auf die unterschiedlichen QUS-Verfahren. Die Ausführungen in dem vorliegenden Bericht beziehen sich auf folgende Teilaspekte:

- Beschreibung des Technologie-Status
- Performance (d. h. Funktion / Messcharakteristika)
- Vorhersagewert für Frakturrisiko – auch im Vergleich zu anderen Technologien
- Technologiegebrauch in Kanada

c) Methodik

Der Bericht gibt einen systematischen Überblick über die verfügbare Literatur unter besonderer Berücksichtigung der analytischen Genauigkeit, der Wirksamkeit und der Effektivitätsbeurteilung.

Es wurde in den elektronischen Datenbanken Medline, EMBASE, HealthStar und Current Contents nach Literatur mit den Suchwörtern ‚calcaneal‘ und ‚ultrasound.ti‘, ‚sh‘, ‚exp Calcaneus‘, ‚exp densitometry‘, ‚exp osteoporosis‘, ‚postmenopausal‘, ‚exp Mass screening‘, ‚exp Bone density‘, ‚ultrasonography‘ in den Jahrgängen 1988 – 1998 gesucht. Insgesamt wurden 84 Literaturstellen ausgewertet, darunter auch Abstracts, die auf dem Jahrestreffen des American College of Rheumatology und der American Society for Bone and Mineral Research präsentiert wurden.

Ein- und Ausschlusskriterien wurden nicht dokumentiert.

d) Ergebnisse, Diskussion und Schlussfolgerungen

Relevante Daten konnten für insgesamt 5 Geräte, die auf dem kanadischen Markt vertreten sind, identifiziert werden, dies sind im Einzelnen: Lunar Achilles, McCue CUBA, Walker Sonix 575, Hologic Sahara, Norland Paris.

Präzision und Richtigkeit

Eine Reihe von Studien liefern Werte über die Kurzzeit-Präzision und den Präzisionsfehler.

Präzisionsangaben fand man in zwei Studien (Strelitzki et al. 1996; Machado et al. 1997), wobei sich in der 1. Studie, in der Phantomkalibrierungen über ein Jahr durchgeführt wurden, bei den BUA-Werten unterschiedlicher Geräte (Lunar Achilles, Walker Sonix, McCue CUBA) ähnliche Prozentwerte zeigten (0.75 % - 1.46 %). Die Präzisionsangaben für die SOS-Messwerte wiesen eine größere Variationsbreite (0.25 % - 2.77 %) auf.

In der 2. Studie, in der Wiederholungsmessungen an osteoporotischen und gesunden Frauen mit drei Geräten durchgeführt wurden (Lunar Achilles, Walker Sonix, CUBA), finden sich ähnliche Präzisionsfehler für die Werte BUA (9.7 % - 13.2 %) und SOS (5.3 % - 9.7 %) Die Übersicht der Daten wird in einer Tabelle dokumentiert, weitere Studiendetails sind im Anhang des kanadischen HTA-Berichtes aufgelistet.

Auch für die „long-term“ Präzision finden sich nur wenig publizierte Zahlen. Lediglich in zwei Studien findet man Angaben über die Langzeit-Präzision, einmal über ein Jahr mit einem Wert von 3.3 % beim *Walker Sonix* (Stewart 1994) und einmal über drei Monate mit einem Wert von 4.9 % für Bua und 1.3 % für SOS beim *Cuba Clinical* (Graafmans et al. 1996).

Angaben zur Richtigkeit (accuracy) sind nach Aussage der Autoren in den vorliegenden Studien nicht verfügbar. Es ist zwar möglich, in vitro physikalische Knocheneigenschaften mit Hilfe der Ultraschallverfahren zu beschreiben und es ist auch möglich zwischen osteoporotischen (mit Frakturen) und "normalen" Individuen zu diskriminieren. Dennoch ist nicht klar, welche Knochenqualität genau von den Ultraschallverfahren erfasst wird. Ein "wahrer" Messwert für QUS-Verfahren lässt sich daher nicht eindeutig definieren.

Als Störfaktoren für die Richtigkeit und Präzision bei QUS-Messungen am Fersenbein werden Rechts- oder Links-Händigkeit, Wasserbad-Temperatur, Sprunggelenksödem, Fettanteil des Gewebes und sogar die Verwendung von Leitungswasser anstatt abgekochtem Wasser angegeben.

Frakturvorhersagefähigkeit

Dieser Teilaspekt wird in insgesamt 8 Studien untersucht: Das Walker Sonix Gerät in drei prospektiven Studien (Bauer et al. 1997; Porter et al. 1990; Stewart et al. 1996), das Lunar Achilles Gerät in zwei Studien (Hans et al. 1996; Thompson et al. 1996), das CUBA clinical Gerät in zwei Studien (Graafmans et al. 1996; Pluijm et al. 1997) sowie das Hologic Sahara Gerät in 1 Querschnittsstudie (Kolta et al. 1997). Darüber hinaus wird das Walker Sonix Gerät in vier Studien einer ROC-Analyse unterzogen, das Lunar Achilles Gerät in drei Studien.

Die meisten Studien zeigen eine gute Korrelation der Ultraschall-Parameter zur Fraktur-Prävalenz der Studienpopulation. Das relative Frakturrisiko (RR) verbunden mit niedrigen Ultraschallmesswerten stellt sich ähnlich den korrespondierenden Daten einer DXA-Messung an der Wirbelsäule und Hüfte dar, die Werte werden einzeln für jedes Gerät in einer Tabelle aufgeführt.

Am ausführlichsten werden für den *Walker Sonix 575* die Ergebnisse hinsichtlich der Fähigkeit zur Frakturvorhersage in drei prospektiven Kohortenstudien diskutiert (Bauer 1997; Porter 1990; Stewart 1996). Der Wert für das Relative Risiko für Hüftfrakturen liegt in der Studie von Bauer bei 2.0 (95 %CI 1.5 - 2.7), für alle Frakturen bei 1.3 (95 % CI 1.2 - 1.5), in der Studie von Stewart et al. bei 1.4 (95 % CI 1.2 - 2.4) für alle Frakturen. In der Studie von Kotzki et al. (1994) wird kein Relatives Risiko angegeben sondern Frakturraten zwischen 1.5 % für hohe BUA-Werte und 12.8 % für niedrige BUA-Werte.

In zwei (Hans 1996; Thompson 1996) von sechs prospektiven Kohortenstudien wird über die Frakturvorhersage des *Lunar Achilles* Gerätes berichtet: RR für Hüftfraktu-

ren SOS 1.9 (95 % CI 1.5 - 2.3) und BUA 2.1 (95 % CI 1.7 - 2.6), in der zweiten Studie wird über den Kombinationsparameter Stiffness-Index berichtet.

Begrenzte Daten sind in nur einer prospektiven Studie (Pluijm 1997) für das *McCue CUBA clinical* Gerät zu finden: bei einem Zwei-Jahres-Follow-Up wurden RR-Werte für Hüftfrakturen von 4.5 (95 % CI 1.1 - 16.5) für BUA gemessen. Daraus wurde geschlossen, dass sich der BUA-Messwert zur Frakturrisikoabschätzung eignet im Gegensatz zum SOS-Messwert.

Für das *Hologic Sahara* Gerät wird keine direkte Evidenz betreffend des Frakturvorhersagewertes dokumentiert (Kolta 1997).

Insgesamt wird in sieben Studien (Bauer 1997; Bauer 1995; Glüer 1996; Gonnelli 1995; Hans 1996; Ross 1995; Turner 1995), in denen die Frakturrisikoabschätzung mit DXA-Werten und Ultraschallmessungen ermittelt wird, eine Analyse mit logistischen Regressionsmodellen durchgeführt und dabei ein statistisch signifikanter Anstieg des Frakturrisikos bei niedrigen US-Werten (BUA) aufgezeigt. Die Spanne des Relativen Risikos (RR) betrug 1.1 – 1.6. Eine Studie gab einen RR-Wert von über 2.0 an, der wahrscheinlich aufgrund der unterschiedlichen Zielgrößen (Wirbelkörperdeformitäten, nicht Hüft- oder Wirbelkörperfrakturen) berechnet wurde (Hans 1996).

Diskussion und Schlussfolgerungen

Die kanadischen Autoren kommen zu dem Schluss, dass das quantitative Ultraschallverfahren (zum Zeitpunkt des HTA-Berichtes) nicht als Screening-Instrument für die Allgemeinbevölkerung (mit nachfolgender Bestätigung durch DXA-Untersuchungen) zur Osteoporosediagnostik geeignet sei; eine kombinierte Strategie würde die Kosten im Gesundheitswesen deutlich erhöhen. Es wurden keine Daten gefunden, die eine verbesserte Fraktur-Vorhersagefähigkeit bei Kombination beider Messverfahren (QUS **und** DXA) belegen.

Aufgrund des Präzisionsfehlers der QUS-Technik müssten Verlaufskontrollen weiterhin mit DXA-Verfahren durchgeführt werden.

Zusammenfassend stellen Homik und Hailey fest, dass das quantitative Ultraschallverfahren (QUS) im Fersenbeinbereich eine vielversprechende diagnostische Technologie sei, deren Rolle in der Osteoporosediagnostik jedoch noch nicht eindeutig definiert ist. Weiterer Forschungsbedarf in bezug auf Langzeit-Präzision, Frakturrisikoabschätzung und potentielle Kosten-Effektivität ist erforderlich zur Klärung des Stellenwertes der QUS-Verfahren in der alltäglichen Gesundheitsversorgung.

e) Abschließende Beurteilung

Der HTA-Bericht enthält eine systematische Literaturübersicht unter Einbezug von Studien bis zum Publikationsjahr 1997. Die Methodik der Literaturübersicht kann als systematisch gelten, obwohl einzelne Aspekte in der Dokumentation fehlen (Qualitätsbewertung der Primärmaterialien). Die Fragestellung der Übersicht entspricht genau der zentralen Fragestellung unseres Kurz-HTA-Vorhabens. Die Tatsache, dass

für den kanadischen HTA-Bericht nur Literatur zu den in Kanada verfügbaren Geräten ausgewertet wurde, stellt die Übertragbarkeit der Ergebnisse nicht in Frage, da die gleichen Verfahren auch auf dem europäischen Markt eingeführt waren.

Unseres Erachtens bilden die Ergebnisse von Homik und Hailey eine valide Basis für die Verfahrensbewertung, die allerdings um die Ergebnisse neuerer Studien ergänzt werden muss. Sowohl die Verfahrenstechnik als auch die Auswertungskonzepte (standardisierte Präzisionsangaben) haben sich seit Redaktionsschluss für die kanadische Publikation erheblich weiterentwickelt.

II C.5.2 Systematische Literaturübersichten

Die in diesem Kurzgutachten angewendete Literaturrecherche nach hierarchischem Vorgehen sieht vor, neben HTA-Berichten auch systematische Literaturzusammenfassungen zu identifizieren. Eine große Anzahl von Reviews (39) wurde gefunden, nach erster Sichtung erfüllte jedoch nur einer die methodischen Anforderungen (sh. Methodenmanual SOP 2 und 3) und wurde der Bewertung mittels Checklisten unterzogen. In der von Prins et al. (1997) publizierten Literaturübersicht *„The Role of Quantitative Ultrasound in the Assessment of Bone“* wird die Fragestellung *„Wie valide lässt sich das Frakturrisiko mithilfe der quantitativen Ultraschallverfahren im Vergleich zum DXA-Verfahren abschätzen?“* anhand der publizierten Literatur bearbeitet. Basis für die Übersicht bildet eine systematische Literaturrecherche in der Medline Datenbank, Jahrestreffen und Abstracts großer Fachgesellschaften sowie Durchsicht der Current Contents bis einschließlich 1. April 1997. Leider fehlen zum weiteren methodischen Prozedere, insbesondere zur Selektion der einzuschließenden Studienergebnisse, sämtliche Angaben, sodass wir aufgrund der methodischen Intransparenz und des Alters der Publikation beschlossen, die Ergebnisse der Arbeit im Rahmen unseres Kurz-HTA nicht ausführlich darzustellen.

II C.5.3 Ergebnisse neuer Primärstudien

In dem vorliegenden HTA-Bericht von Homik und Hailey wurden nur Studien mit Publikationsdatum bis einschließlich 1997 berücksichtigt. Es stellte sich daher die Frage, ob neuere Studien Ergebnisse liefern, die die Schlussfolgerungen des HTA-Berichtes bzw. die Beantwortung der für dieses Kurz-HTA relevanten Fragestellungen modifizieren könnten. Die Literaturanalyse bis 1997 zeigte, dass die quantitative Ultraschalldiagnostik zu jenem Zeitpunkt keine ausreichend valide und zuverlässige Abschätzung des Frakturrisikos erlaubte (Homik & Hailey 1998). Unsere ergänzende Recherche über die Publikationszeiträume 1998 - 2001 wurde daher so konzipiert, ausschließlich prospektive Kohortenstudien aufzufinden, deren Zielgröße die Vorhersagefähigkeit der QUS-Verfahren für osteoporotische Frakturen war.

Die Datenbanken und Recherchestrategien sowie die Ein- und Ausschlusskriterien befinden sich im Anhang.

Insgesamt wurden sieben prospektive Studien zur Thematik aufgefunden, davon wurden zwei Studien aus methodischen Gründen ausgeschlossen.

- Do markers of bone resorption add to bone mineral density and ultrasonographic heel measurement for the prediction of hip fracture in elderly women? (Garnero et al., EPIDOS, Osteoporosis Int 8: 563-569; 1998), da es sich nicht um eine prospektive Studie, sondern um eine eingestrichelte Fall-Kontroll-Analyse handelte und
- Possible predisposing factors for the second hip fracture (Dretakis et al., Calcified Tissue Int 62: 366-369; 1998). Die Informationen zum prospektiven Teil lassen sich nur in der Diskussion finden, sodass ebenfalls kein prospektives Studiendesign vorliegt. Es handelt sich um eine Fall-Kontroll-Studie.

Die Einschlusskriterien (Ultraschallmessung, Frakturrisiko, prospektiv) wurden nur von fünf Studien erfüllt, deren Charakteristika und Hauptergebnisse im folgenden ausführlich dargestellt werden. Die tabellarische Dokumentation findet sich im Anhang. In den USA wurde eine Arbeit 1998, je eine Arbeit in Großbritannien, Frankreich und den Niederlanden im Jahr 1999 und eine italienische Arbeit im Jahr 2000 publiziert.

II C.5.3.1 Gnudi et al.: Quantitative Ultrasound and Bone Densitometry to Evaluate the Risk of Nonspine Fractures: A Prospective Study; 2000

Fragestellung: Die Fragestellung dieser aktuellen prospektiven Studie aus Italien zielt auf die Fähigkeit der QUS-Verfahren ab, das Risiko für osteoporotische Frakturen vorherzusagen sowie auf die Beurteilung, ob kombinierte Messungen von QUS und BMD eine bessere Abschätzung des Frakturrisikos bei Osteoporose-Patienten ergeben als bei Anwendung von nur einem Verfahren.

Studiendesign: Prospektive Kohortenstudie

Studienpopulation: 254 zufällig ausgewählte postmenopausale Frauen (im Mittel 58 Jahre), die an einem Knochendichte-Screening in einem Osteoporose-Zentrum teilnahmen. Ausgeschlossen wurden Patientinnen mit Tumorerkrankungen, Erkrankungen aus dem rheumatischem Formenkreis, Paget-Erkrankung, chronischen Nieren-, Leber- oder Lungenerkrankungen sowie endokrinen Erkrankungen, Langzeitimmobilisation und Hüft- oder Knieprothese. Eingeschlossen wurden auch Frauen, die eine medikamentöse Osteoporosetherapie (z. B. Calcitonin, Östrogene, Bisphosphonate, Vitamin D und Calcium-Substitution) erhielten.

Expositionsmessung: QUS: Das quantitative Ultraschallmessverfahren wurde am distalen Radius (DR) und an der Patella (P) mit dem Signet-Gerät (Osteotechnology Inc., Framingham, MA) durchgeführt. Als Messgröße wurde die UTV (ultrasound transmission velocity) an beiden Messorten ermittelt.

BMD: Die BMD-Messungen wurden nur am distalen Radius (DR) mit einem Single-energy-densitometer (Norland 2780, Norland Corp, Fort Atkinson, WI) durchgeführt. Die Kalibrierung erfolgte täglich mit einem gerätespezifischen Phantom. Der populationspezifische T-Score wurde in einer Gruppe von 64 jungen Frauen ermittelt.

Zielgrößen: Frakturen (nicht Wirbelkörper) wurden in jährlichen Abständen durch klinische und/oder radiologische Untersuchungen dokumentiert, wobei nur Knochenbrüche als Folge von Minimaltraumata berücksichtigt wurden. Der Nachuntersuchungszeitraum betrug im Mittel 5.47 Jahre (± 1.05) mit einer durchschnittlichen Nachkontrollzeit vom Zeitpunkt des Beobachtungsbeginns bis zum Auftreten der Erstfraktur von 2.72 Jahren (± 1.47).

Ergebnisse: In den jährlichen Untersuchungen wurden insgesamt 50 Frakturen entdeckt. Auswertungen im Cox-Regressionsmodell ergaben folgende relative Frakturrisiken für eine Standardabweichung (SD) vom Mittelwert (adjustiert für Alter, Menopausenalter, Größe, Gewicht, Behandlung):

BMD-Dr	2.99	(1.06 – 8.41)	für osteoporotische Frakturen
UTV-Dr	3.69	(1.18 – 11.49)	für osteoporotische Frakturen
UTV-P	3.89	(1.53 – 9.90)	für osteoporotische Frakturen

(UTV = *Ultrasound transmission velocity*, DR = *Distaler Radius*, P = *Patella*)

Zur Analyse der Diskriminationsfähigkeit der Verfahren zwischen gesunden und frakturgefährdeten Frauen wurden ROC-Analysen durchgeführt.

Diskussion und Schlussfolgerung der Autoren: Der Vergleich der Verfahren QUS und DXA in ROC-Analysen konnte keine unterschiedliche Vorhersagefähigkeit für osteoporotische Frakturen feststellen. Die Autoren betonen, dass der BMD-Dr und der UTV-Dr-Wert eher ein erhöhtes Risiko für Handgelenksfrakturen und der UTV-P-Wert eher Hüft- und Knöchelfrakturen vorhersagt.

In der Diskussion und Beurteilung wird von den Verfassern betont, dass bezogen auf die gesamte Untersuchungspopulation, der Abfall der QUS-Messwerte (UTV-Dr und UTV-P) signifikant mit dem Anstieg der nicht-spinalen Frakturen korreliert. Das relative Risiko (RR) eine osteoporotische Fraktur zu erleiden reicht von 4.49 bis 5.35. Dagegen fanden sich keine statistisch signifikanten Unterschiede zwischen den RR-Werten der Knochendichtemessung mittels DXA und denen der hier untersuchten QUS-Verfahren. Diese Daten stimmen mit denen aus der internationalen Literatur (Nevitt 1994; Wasnich 1989; Duppe 1997; Hans 1996; Bauer 1997; Mele 1997) überein, beide Verfahren können demnach das Frakturrisiko bei osteoporotischen Frauen vorhersagen.

Innerhalb der Untergruppe der unbehandelten, gesunden Frauen (125) liessen sich keine signifikanten Korrelationen zwischen den Messwerten und dem Frakturrisiko finden.

Vorausgegangene stattgehabte Frakturen haben einen sehr hohen Frakturvorhersagewert, bei handgelenksnahen ist dieser sogar höher als die Kombination des QUS-Verfahrens und der BMD-Messungen. Die gewonnenen Daten unterstützen die Theorie, dass das Frakturvorhersagevermögen vom Messort und Frakturtyp abhängt und dass die Knochenzusammensetzung eine große Rolle bei der Abschätzung des Frakturrisikos mit Hilfe der quantitativen Ultraschalldiagnostik spielt.

Abschließend muss hervorgehoben werden, dass die in der Studie verwendeten Geräte der Ultraschalltechnik heute obsolet sind. Darüberhinaus ist anzumerken, dass eine nicht homogene Studienpopulation untersucht wurde, da auch bereits therapierte Osteoporosepatientinnen teilnahmen.

II C.5.3.2 Stewart et al.: Predicting a Second Hip Fracture; 1999

Fragestellung: Das Ziel dieser prospektiven Studie aus Großbritannien war die Identifikation von Risikofaktoren für Zweitfrakturen der Hüfte, mit dem Ziel, Empfehlungen für eine entsprechende Prophylaxe abzuleiten

Studiendesign: prospektive Kohortenstudie

Studienpopulation: 3119 Patienten über 50 Jahre, die in einem definierten Zeitraum (1987 - 93) eine Schenkelhalsfraktur erlitten hatten, wurden zu dieser Studie eingeladen. Aus 1467 Überlebenden der Gesamtpopulation stimmten 394 Personen einer Studienteilnahme zu.

Expositionsmessung: QUS: Das quantitative Ultraschallmessverfahren wurde am Fersenbein mit zwei Geräten durchgeführt: a) trockenes System McCue CUBA Clinical, bei welchem die Messgrößen BUA (McC) und VOS (velocity of sound) ermittelt wurden sowie b) feuchtes System Walker Sonix UBA 575, bei welchem die Messgröße BUAWs bestimmt wurde. Die Präzision der QUS-Messungen (Phantomkalibrierung) war 2.6 % für BUAWs, 1.3 % für VOS, 5.4 % für BUAMcC.

BMD: Die BMD-Messungen wurden an drei Körperregionen durchgeführt: Schenkelhals, Trochanter major, Ward-Dreieck, verwendet wurde das DXA-Verfahren (Norland XR-26 scanner Norland Corp., WI, USA). Die Präzision dieser Geräte wurde wie folgt angegeben: 2.8 % für Schenkelhals, 1.9 % für Trochanterregion und 4.9% für Ward-Dreieck.

Weitere Untersuchungen: Mini Mental State Examination (MMSE), National Adult Reading Test (NART), Hospital Anxiety and Depression Score, Gibson Spiral Maze score, Barthel Index, Nottingham ADL sowie Sehtest nach Snellen.

Zielgröße: Zielgrößen wurden im Verlauf der Studie in 6 - 12 monatigen Intervallen [im Mittel 51.8 Wochen (25-80)] und in einer Abschlussuntersuchung nach vier Jahren Gesamtbeobachtungsdauer erhoben. Während des Follow-Up wurden Frakturen, Stürze und Lebensgewohnheiten mit einem postalischem Fragebogen erfasst, Frakturen wurden anhand von angeforderten Röntgenaufnahmen bzw. anhand von Kran-

kenhausakten verifiziert. In der Abschlussuntersuchung wurde die primäre Zielgröße anhand der Krankenhausentlassungsstatistik für Responder und für Nicht-Responder ermittelt. (Es wurden Zweitfrakturen der Hüfte identifiziert, 27 Patienten erlitten eine solche während des Follow-up).

Ergebnisse: 27 Patienten aus der Studiengruppe (Responder) erlitten eine Refraktur der Hüfte, die Rate betrug damit 6.85 % für den postoperativen Zeitraum von 5 – 10 Jahren. Bei den Non-Respondern betrug die Refrakturrate 5.87 %. Relative Refrakturrisiken (pro 1 SD Abweichung vom Mittelwert) wurden in Abhängigkeit von Ultraschallparametern, Knochendichteparametern, Mobilität, Stürzen, psychischen Auffälligkeiten ermittelt (adjustiert für Alter, Geschlecht und Körpergewicht). Für die technischen Messverfahren ergaben sich dabei folgende Werte:

Tabelle 30: RR für Refraktur in Abhängigkeit von QUS und BMD; nach Stewart et al. 1999

Parameter	RR für - 1 SD vom Mittelwert	95 % Konfidenzintervall
BUA McC	1.22	(0.74-2.02)
SOS	1.36	(0.81-2.29)
BUA ws	1.17	(0.72-1.91)
SH-BMD	1.64	(0.92-2.91)
TB-BMD(TB=Total Body)	1.15	(0.67-1.98)

Tabelle 31: ROC-Analyse QUS und BMD-Messungen; nach Stewart et al. 1999; ^a signifikant ($p < 0,05$) kleiner als Trochanter BMD

Parameter	Area Under ROC-Curve (AUC)	95 % Konfidenzintervall
BUA McC ^a	0.596	SE 0.055
SOS ^a	0.537	SE 0.060
BUA ws ^a	0.562	SE 0.055
SH-BMD	0.699	SE 0.046
Trochanter BMD	0.705	SE 0.046

In der ROC-Analyse erweisen sich die DXA-Messwerte am Schenkelhals und Trochanter major als signifikant bessere Vorhersagefaktoren als die QUS-Parameter.

Diskussion und Schlussfolgerungen der Autoren: Obwohl in verschiedenen anderen Studien bewiesen wurde, dass die Knochendichte Frakturen voraussagen kann, insbesondere bei Messungen an der Hüfte (Cummings 1995) aber auch bei QUS-Messungen am Kalkaneus (Hans 1996; Bauer 1997) kommen die Autoren zu dem Schluss, dass die Vorhersage eines Refrakturrisikos nur durch Addition weiterer Risikoindikatoren erreicht werden kann. Sie gehen sogar so weit in ihrer Beurteilung, dass sie Knochendichtemessungen zur Ermittlung eines Refrakturrisikos für über-

flüssig erachten, da z. B. die (prophylaktisch angehbare) Kombination aus niedrigem Körpergewicht und Bewegungsarmut ausreicht, um eine Gefährdung zu erkennen.

Allerdings muß bei dieser Studie beachtet werden, dass es sich um die Spezialsituation „2. Hüftfraktur“ handelt und somit die Studienpopulation eine ältere Hochrisikogruppe ist.

II C.5.3.3 Dargent-Molina et al.: Separate and Combined Value of Bone Mass and Gait Speed Measurements in Screening for Hip Fracture Risk: Results from the EPIDOS Study; 1999

Fragestellung: Ziel der vorliegenden Auswertung, die auf Daten der EPIDOS-Studie (Epidemiologie de l'Osteoporose) basiert, war es, die prädiktive Eigenschaft von SH-BMD (Knochendichte am Schenkelhals), BUA am Kalkaneus, Ganggeschwindigkeit und Alter für das Hüftfrakturrisiko bei älteren Frauen zu untersuchen. Darüber hinaus sollte festgestellt werden, ob die Kombination von zwei oder mehreren dieser Messungen eine Verbesserung der Vorhersagefähigkeit erreichen kann.

Studienpopulation: Insgesamt wurden in der EPIDOS-Studie, die von Januar 1992 bis Januar 1994 in fünf französischen Großstädten durchgeführt wurde, 7.575 Frauen im Alter über 75 Jahre (im Mittel 80.5 ± 3.8 Jahre) prospektiv auf Risikofaktoren für Hüftfrakturen untersucht. Da nur bei einem Teil der Kohorte eine QUS-Untersuchung durchgeführt wurde, reduziert sich die Anzahl der in dieser Studie ausgewerteten Teilnehmerinnen auf 5.978.

Expositionsmessung: QUS: Die quantitative Ultraschallmessung wurde am Fersebein mit dem Lunar Achilles Device (Lunar Corporation, Madison, WI) durchgeführt.

BMD: Die BMD-Messung am Schenkelhals erfolgte mit Hilfe des DXA-Verfahrens (Lunar DPX-Plus).

Gang: Die normale Ganggeschwindigkeit wurde auf einer 6-Meter-Strecke als Mittelwert aus zwei Untersuchungen ermittelt.

Zielgrößen: Zielgröße war die Hüftfraktur, deren Erstauftreten in viermonatigen Intervallen per Fragebogen erfasst wurde.

Ergebnisse: Die Responserate betrug 96.3 %. 206 Frauen verweigerten die weitere Teilnahme, 15 Frauen waren nicht mehr auffindbar. In dem im Mittel 33-monatigen Nachuntersuchungszeitraum (± 8.6 Monate) wurden insgesamt 170 Hüftfrakturen registriert, 340 Frauen verstarben.

Die Eignung der Messverfahren zur Erkennung von Risikopatientinnen für Hüftfrakturen wurde anhand der Sensitivität bei unterschiedlichen zugrundegelegten Cut-off-Levels ermittelt (Tabelle 32 und Tabelle 33).

Tabelle 32: Sensitivitäten von BMD, QUS, Gang und Alter für Hüftfrakturen in Abhängigkeit vom Cut-off (nach Dargent-Molina 1999)

Cut-off	Sensitivität (%) von			
	BMD-SH	BUA Kalkaneus	Ganggeschwindigkeit	Alter
50. Perzentile	78	69	69	68
25. Perzentile	49	43	42	42
10. Perzentile	25	25	24	26

Tabelle 33: Sensitivitäten von Kombination aus BMD, QUS, Gang und Alter für Hüftfrakturen in Abhängigkeit vom Cut-off (nach Dargent-Molina 1999)

Cut-off	Sensitivität (%) von Alter +						
	BMD-SH	BUA (Kalk)	Gang	BMD + BUA	BMD + Gang	BUA + Gang	BMD + BUA + Gang
50. Perzentile	85	74	79	85	82	80	85
25. Perzentile	52	55	49	53	59	59	59
10. Perzentile	29	29%	28	31	32	34	36

Die ermittelten Werte machen deutlich, dass die Sensitivitäten aller Messverfahren relativ niedrig sind, wobei Knochendichtemessungen am Schenkelhals, die BUA-Messung am Kalkaneus, die Ganggeschwindigkeit und das Alter größenordnungsmäßig die gleiche Sensitivität zur Identifikation von Hüftfrakturpatientinnen aufweisen. Um Sensitivitäten von über 80 % zu erhalten, müssen Parameter kombiniert und jeweils die 50. Perzentile als Cut-Off gewählt werden. Damit wären 50 % der Bevölkerung als Risikopatienten für Hüftfrakturen anzusehen.

Das für Gang und Alter adjustierte Relative Risiko (RR) für Hüftfrakturen wurde in dieser Studie so angegeben:

BMD (*Lunar DPX-plus*) femoral (-1SD) 1.6 95 % KI (1.38 – 1.95)

BUA (*Lunar Achilles*) calcaneal (-1SD) 1.4 95 % KI (1.2 – 1.7)

Diskussion und Schlussfolgerungen durch die Autoren: Die vorliegenden Ergebnisse werden von den Autoren wie folgt kommentiert:

Einbezug des Alters in einen Risikoscore verbessert die Sensitivität aller Einzelparameter; da diese Information eigentlich immer vorhanden ist, sollte sie systematisch in die Risikobestimmung integriert werden. Zur Erkennung einer Hochrisikogruppe für Hüftfrakturen könnte anstelle der BMD-SH Messung das quantitative Ultraschallverfahren herangezogen werden. Bei den Cut-offs von 25 % und 10 % sind die Sensi-

vitäten der Verfahren vergleichbar, wobei das Ultraschallverfahren kostengünstiger ist und keine Strahlenexposition mit sich bringt.

Die einfache, schnelle und kostengünstige Messung der Ganggeschwindigkeit hat den gleichen Diskriminanz-Wert wie die BMD-Messung am Schenkelhals. Orientiert am möglichen Therapieansatz (z. B. Krankengymnastik oder medikamentöser Knochenaufbau) sollte auch die Screening-Methode (Ganggeschwindigkeit oder BMD-Messungen) ausgewählt werden.

Abschließend stellen die Verfasser die Frage zur Diskussion, ob zur signifikanten Reduktion der Hüftfrakturinzidenz bei Frauen über 75 Jahre eine "Massenscreening"strategie (Vorsorgeuntersuchung für alle und Behandlung der Hochrisikopatienten) sinnvoll erscheint, zumal die Sensitivität der zur Verfügung stehenden Screeninginstrumente auch in der Kombination niedrig ist.

II C.5.3.4 Pluijm et al.: Ultrasound Measurements for the Prediction of Osteoporotic Fractures in Elderly People; 1999

Studienziel: Das Studienziel dieser aus den Niederlanden kommenden prospektiven Studie war die Untersuchung des Vorhersagewertes für das Frakturrisiko mithilfe der quantitativen Ultraschall-Messungen (trockenes System) sowie der Vergleich des prädiktiven Wertes von QUS-Parametern mit anderen potentiellen Frakturrisikofaktoren in der älteren Bevölkerung.

Studienpopulation: Die Studienpopulation setzt sich aus 710 Altenheimbewohnern über 70 Jahre (im Mittel 82.8 ± 5.9) ohne wesentliche kognitive Störungen zusammen, darunter 132 Männer und 578 Frauen.

Expositionsmessungen: QUS: BUA und SOS wurden mit dem „trockenen System“ *Cuba clinical instrument* (McCue Ultrasonics, Winchester, UK) je zwei mal an der rechten und linken Ferse bei jedem Teilnehmer gemessen. Die Präzision wird als Variationskoeffizient mit 3.4 % für BUA und 1.3 % für SOS angegeben.

Körpergewicht

Strukturiertes Interview: Demographie, Frakturanamnese, Gebrauch von Gehhilfen, physischer Aktivitätslevel, Mobilität, Stürze.

Zielgrößen: Sämtliche erlittenen Frakturen während eines Follow-up von im Mittel 2.8 Jahren (max. 3,7 Jahren) wurden mittels halbjährlicher Telefonabfrage oder im postalischen Fragebogen von den Teilnehmern erfragt. Pflegepersonal wurde interviewt, wenn der Teilnehmer nicht in der Lage war selber zu antworten. Von im Intervall verstorbenen Teilnehmern wurden Frakturereignisse von den Hausärzten erfragt. Jede berichtete Hüftfraktur wurde ebenfalls anhand der Unterlagen des Hausarztes verifiziert.

Ergebnisse: 168 (23.7 %) Patienten verstarben innerhalb des Studienzeitraums, 5 (0.7 %) waren nicht mehr erreichbar. Insgesamt wurden 96 nichtspinale Frakturen, 31 Hüftfrakturen und 65 weitere Frakturen wie Colles- oder Oberarm-Brüche dokumentiert.

Zur Schätzung des Zusammenhanges von QUS-Parametern und Frakturen, sowie zur Identifikation weiterer möglicher Prädiktoren wurde eine Cox-Regressions-Analyse herangezogen.

Das für Alter und Geschlecht adjustierte Relative Risiko (RR) eine Fraktur zu erleiden wird für 1 Standardabweichung (-1 SD) für beide QUS-Parameter (*Cuba clinical*) wie folgt angegeben:

Tabelle 34: Vorhersagewerte (RR) für BUA und SOS (-1 SD Differenz) für Frakturen unterschiedlicher Lokalisation (alters- und geschlechtsadjustiert) (nach Pluijm et al. 1999)

Parameter	Frakturort	RR (95 % KI)
BUA	Hüfte	2.3 (1.4 – 3.7)
	andere, nicht spinale Frakturen	1.3 (0.9 – 1.8)
	Frakturen insgesamt	1.6 (1.2 – 2.1)
SOS	Hüfte	1.6 (1.1 – 2.3)
	andere, nicht spinale Frakturen	1.1 (0.9 – 1.5)
	Frakturen insgesamt	1.3 (1.0 – 1.6)

Beide QUS-Parameter sind assoziiert mit ansteigendem Risiko für Hüftfrakturen und die Gesamtheit aller Frakturen, nicht aber für „nicht-spinale“ Frakturen. Unter den demographischen und klinischen Faktoren waren Alter, Immobilität > vier Wochen im letzten Jahr (Hüft-Fx, alle Fx), Körpergewicht < 67 kg (Hüft-Fx, nicht-spinale Fx, alle Fx) und weibliches Geschlecht (nicht-spinale Fx, alle Fx) signifikant mit erhöhten Frakturrisiken verbunden.

Diskussion und Schlussfolgerungen: Die Autoren schliessen aus den Ergebnissen, dass niedrige BUA- und SOS-Messungen das Hüftfrakturrisiko und das allgemeine Frakturrisiko bei älteren Menschen vorhersagen können. Der BUA-Parameter wurde als der bessere Vorhersagefaktor identifiziert.

Diese Resultate untermauern die Feststellungen von Fall-Kontroll-Studien (Baran 1988; Heaney 1989; Bauer 1995; Schott 1995; Dretakis 1995; Turner 1995; Glüer 1996), dass quantitative Ultraschallmessungen zwischen Fraktur- und Kontrollgruppen unterscheiden können und bestätigen die Ergebnisse der bekannten prospektiven Studien (Hans 1996; Bauer 1997).

Bei der Interpretation der Ergebnisse geben die Autoren zu bedenken, dass die Inzidenz einer ersten Hüftfraktur und die Anzahl der teilnehmenden männlichen Probanden in dieser Studie sehr klein war, sodass die Schätzungen für diese Gruppe mit

einer erheblichen Unsicherheit behaftet sind; darüber hinaus war die Auswahl der Studienpopulation eher eng gesteckt (fortgeschrittenes Alter, institutionalisiert) und die Frakturangaben basierten nur auf Eigenberichten. Somit kann eine Unterschätzung des Frakturrisikos entstehen.

II C.5.3.5 Huang et al.: Prediction of Fracture Risk by Radiographic Absorptiometry and Quantitative Ultrasound: A Prospective Study; 1998

Studienziel: In dieser bevölkerungsbezogenen Studie, der sogenannten Hawaii Osteoporosis Study (HOS), wurde die Fähigkeit zur Frakturrisikovorhersage mittels Computer-assisted Radiographic Absorptiometry (RA) und quantitativen Ultraschallverfahren untersucht.

Studienpopulation: Hierzu wurden Longitudinaldaten aus den HOS-Untersuchungen 8 (1992 - 94) und 9 (1995 - 96) herangezogen, die an 560 postmenopausalen japanisch-amerikanischen Frauen mit einem Durchschnittsalter von 74 (55 - 92) Jahren und einem Nachuntersuchungszeitraum von im Mittel 2.7 Jahren erhoben wurden (Zur detaillierten Beschreibung der HOS-Kohorte vergl. Heilbrun et al. 1985).

Expositionsmessung: QUS: Die quantitative Ultraschallmessung wurde mit dem Walker Sonix Model UBA 575 (Hologic Inc, Waltham, MA, USA) durchgeführt. Die BUA-Werte wurden am Fersenbein ermittelt. Die Präzision betrug ausgedrückt als Varianzkoeffizient (CV) 5.8 % calcaneal, 2.2 % metacarpal und 1.2 % phalangeal.

BMD: Zur BMD-Messung kommen zwei Modelle der Computer-assisted Radiographic Absorptiometry (RA) zum Einsatz und zwar a) Osteo Gram (CompuMed, Inc., Manhattan Beach, CA, USA) an den Phalangen und b) Bonealyzer (Model THE-40, Teijin Ltd., Tokyo, Japan) an den Metacarpalia.

Zielgrößen: Als Zielgrößen wurden in dieser Studie inzidente Frakturen analysiert und zwar: vertebrale Frakturen (ja / nein), nicht-spinale Frakturen (ja / nein), alle Frakturen (ja / nein). Eine Wirbelfraktur wird als Höhenverlust des Wirbelkörpers von 15 % im herkömmlichen Röntgenbild definiert. Die Patientenangaben zu Datum, Frakturort und radiographische Bestätigung wurden anhand von medizinischen Berichten überprüft.

Ergebnisse: In einem Cox Regressions Modell wird der Zusammenhang der Frakturrinzidenz mit kalkanealen BUA-Werten als Odds ratios (OR) berechnet:

Wirbelkörperfrakturen	OR 1.50	95% KI	1.05 – 2.16
Nicht-spinale Frakturen	OR 1.89	95% KI	1.27 – 2.88
Frakturen gesamt	OR 1.72	95% KI	1.30 – 2.31

Die Odds ratios sind für – 1 SD kalkuliert, adjustiert für Alter.

Die Frakturwahrscheinlichkeit steigt mit zunehmendem Alter und absteigenden BUA-Werten an.

Diskussion und Schlussfolgerung durch die Autoren: Die vorliegende Studie ist die erste bevölkerungsbasierte Untersuchung, die sich mit der Frakturvorhersagefähigkeit der quantitativen Ultraschallverfahren in einer asiatischen Population beschäftigt.

Auch in dieser Veröffentlichung kommen die Autoren zu dem Schluss, dass der am Fersenbein gemessene BUA-Wert das Frakturrisiko sowohl an der Wirbelsäule als auch an anderen Stellen voraussagt, allerdings scheint eine engere Assoziation der QUS-Werte mit nicht-spinalen Frakturen (OR = 1.89) als mit Wirbelkörper-Frakturen (OR = 1.50) vorzuliegen.

Die Aussagen müssen kritisch betrachtet werden, da die angewendeten Regressionsmodelle auf eine relativ kleine Anzahl von Frakturfällen aufgebaut sind, die in einem kurzen Follow-up beobachtet wurden. Daher steht die Forderung nach einem ausreichend langen Nachuntersuchungszeitraum, um zu valideren Ergebnissen zu kommen.

Nachfolgend sind die Ergebnisse der fünf auswertbaren prospektiven Kohortenstudien noch einmal zusammenfassend in tabellarischer Form dargestellt.

Tabelle 35: Eckdaten der eingeschlossenen Studien

Studie	Methodik	Studienpopulation	Expositionsmessung	Zielgrößen	Ergebnisse	Bemerkungen
Pluijm SMF. 1999	Prospektive Studie 1-Jahres Follow-up Questionnaire	Niederlande 710 Teilnehmer; 132 ml. 578 wbl. ξ 82.8 ± 5.9 Jahre Altersheime Ø Pflegestationen Einschluss: ≥ 70 Jahre Ø kognitive Störung Follow up: ξ 2.8 Jahre 23.7 % (168) Verstorben 0.7 % (5) Verloren	QUS: CUBA Clinical Instrument (McCue Ultrasonics, Winchester/UK) "Trockenes System" Silicongel, Ferse 2 x re. + li. Ferse Präzision: CV 3.4 % BUA CV 1.3 % SOS (20 Freiwillige)	Hüft-Fx.: 31 (von 30 Pat.) Nicht-spinale Fx 96 (77 Pat.) andere nicht-spinale Fx.: 65 (54 Pat.) (23 Colles-Fx., 8 OA-Fx.)	Cox Modell; RR 95% CI Hüft-Fx.: BUA RR 2.3 (1.4-3.7) SOS RR 1.6 (1.1-2.3) Alle Frakturen BUA RR 1.6 (1.2-2.1) SOS RR 1.3 (1.0-1.6) adjustiert für Ge- schlecht und Alter Regressionsanalyse: Hüft-Fx.: BUA RR 2.1 (1.3-3.3) SOS RR 1.4 (1.0-2.1) Jede Fraktur: BUA RR 1.6 (1.2-2.2)	Cave: Population: Alt, Heimbewohner
Huang C. 1998	Hawaii Osteoporosis Study Prospektive Kohorten- studie (longitudinale Daten aus HOS 8 + 9) Bevölkerungsstudie Follow-up: ≈ 2.7 Jahre	Hawai, USA 560 wbl. 66 mit Fx. 494 ohne 251 wbl. 34 mit Fx. 217 ohne 509 wbl. 63 mit Fx. 446 ohne postmenopausale Japanisch-Amerikanisch ≈ 74 Jahre (55-92)	BMD Computer-assisted radiographic absorptiometry (RA) OsteoGram (Compu Med, USA) Phalangen Boneanalyzer (The-40, Teijin Ltd.Japan) QUS Walker Sonix Modell UBA 575 (Halogic Inc. Waltham, Ma, USA BUA – Calcaneus CV 2.2 % metacarpal CV 1.2 % phalangeal CV 5.8 % calcaneal	Wirbelkörper-Fx Nicht WS-Fx Alle Erst-Fx.	BUA calcaneus: Odds ratio 1.5 (1.05-2.16) Odds ratio 1.89 (1.27-2.88) Odds-ratio 1.72 (1.30-2.31) adjustiert für Alter	RA Phalangen QUS Kalkaneus Subgruppenuntersuchung : follow-up Rate

Studie	Methodik	Studienpopulation	Expositionsmessung	Zielgrößen	Ergebnisse	Bemerkungen
Stewart A. 1999	<p>Prospektive Kohortenstudie Fragebogen zw. 6-10 Mon. Follow-up: ≈ 51.8 Wochen (25-80)</p> <p>Präzision: Gesunde wbl. SH 2.8 % Troch 1.9 % Ward 4.9 % BMD 1.1 % BMC 1.0 % BUAws 2.6 % VOS 1.3 % BUAMc 5.4%</p>	<p>n = 394 (SHFx. zw. 1987 - 1993) (gesamt: 2813) > 50 Jahre 64 ml. (16.2 %) 330 wbl. (83,8 %) = 14 % der Hüftfx-Pat. n = 360 (91.4 %) responder 3.3 % verstorben 2 Pat. verzogen 4.8 % ∅ Antwort Ausschluss: 29 Pat. w/ Zweitfraktur</p>	<p>DXA Norland XR-26 Scanner (Norland Corp., WI, USA) Gesunde Hüfte: Schenkelhals Trochanter Ward-Dreieck</p> <p>QUS 1) Walker Sonix UBA575 (BUA ws) 2) McCue CUBA Clinical(BUA McC/ VOS)</p>	<p>Zweit-Fraktur Hüfte 27 Teilnehmer Refrakturrate: 6.85 %</p> <p>Kontrollgruppe: 63 Teilnehmer Refrakturrate: 5.87 %</p>	<p>RR pro 1SD: BUA ws 1.17 (0.72-1.91) SOS 1.36 (0.81-2.29) BUA McC 1.22 (0.74-2.02) SH-BMD 1.64 (0.92-2.91) Total body BMD 1.15 (0.67-1.98)</p> <p>adjustiert für Alter, Geschlecht, Gewicht</p> <p>ROC-Analyse: BUA McC (SE 0.055) AUC 0.596 SOS (SE 0.060) AUC 0.537 BUA ws (SE 0.055) AUC 0.562 SH-BMD (SE 0.046) AUC 0.699 Trochanter BMD (SE 0.046) AUC 0.705</p>	<p>Identifikation von Risikofaktoren für Zweit-Hüft-Frakturen mit dem Ziel einer gezielten Prophylaxe</p>
Dargent-Molina P. 1999	<p>EPIDOS Prospektive Kohortenstudie</p> <p>Fragebogen alle 4 Monate zur Fx-Abfrage</p> <p>Follow-up: 33 Monate (± 8.6 Monate)</p>	<p>Frankreich 5895 wbl. > 75 Jahre (80.5 ± 3.8)</p> <p>96.3 % Teilnehmer 206 Verweigerer 15 Verloren</p> <p>Follow-up: 340 wbl. Verstorben 170 Hüft-Fx.</p>	<p>1) DXA (Lunar DPX-Plus) 2) QUS (Lunar Achilles device , WI,USA) 3) Ganggeschwindigkeit</p>	<p>Hüft-Fraktur (170 wbl.)</p>	<p>Adjusted Relat. Risk (RR): Hüftfrakturrisiko: BUA 1.4 (1.2 - 1.7) BMD 1.6 (1.38 - 1.95)</p> <p>BUA 74 % Sensitivität bei 50 % der Frauen in der Hochrisikogruppe BMD 85 %</p> <p>Adjustiert für Alter</p>	<p>Sensitivitäten von ≈ 70 % nur bei Cut-Off von 50 %</p>

Studie	Methodik	Studienpopulation	Expositionsmessung	Zielgrößen	Ergebnisse	Bemerkungen
Gnudi S. 2000	<p>Prospektive Kohortenstudie</p> <p>Randomisierte Teilnehmer</p> <p>Lineare Regressions-Analyse</p>	<p>Italien</p> <p>254 wbl. Postmenopausal</p> <p>Einschluss: Osteoporosebehand.</p> <p>Ausschluss:</p> <p>Chron. Erkrankungen, Immobilisation, Hüft- oder Knieprothese, Malignom, Rheuma</p> <p>T-Score Kalkulation:</p> <p>64 wbl.</p> <p>36.39 ± 8.9 Jahre</p> <p>Follow-up:</p> <p>5.47 Jahre ± 1.05</p>	<p>Single-energy densitometer (Norland 2780, Norland Corp. Fort Atkinson):</p> <p>Distaler Radius</p> <p>Signet device (Osteotechnology Inc., Framingham)</p> <p>distaler Radius (DR)</p> <p>Patella (P) UTV m/sec</p> <p>Kombinierte Messungen: BMD+QUS</p>	<p>Erfassung der nicht-spinalen Fx.:</p> <p>Jährliche Kontrolle</p> <p>Nachkontrollzeit vom Zeitpunkt des Beobachtungsbeginns bis zum Auftreten der 1.Fraktur: ± 1.47 Jahre</p>	<p>50 Frakturen (3.59/Jahr)</p> <p>Frakturrisiko einzeln: RR (95% KI)</p> <p>BMD-Dr 2.99 (1.06 - 8.41)</p> <p>UTV-DR 3,69 (1.18 - 11.49)</p> <p>UTV-P 3.89 (1.53 - 9.90)</p> <p>adj.f. Alter, Menopausenalter, Größe, Gewicht, Behandlung</p> <p>Frakturrisiko kombiniert: RR (95 % KI)</p> <p>BMD-DR + UTV-DR 13.46</p> <p>BMD-DR + UTV-P 16.93</p> <p>Unbehandelte Studienteilnehmer: ∅ Signifik. zw. Fx + Meßparametern</p>	<p>Geräte nicht mehr im Einsatz</p> <p>Fx-Vorhersage hängt ab von: Meßstelle + Frakturtyp</p> <p>Cave: Technikstudie</p>

II C.5.4 Ergebniszusammenfassung

In vorstehender Tabelle sind die wichtigsten Charakteristika aller fünf im angegebenen Recherchezeitraum identifizierten Publikationen dargestellt. Vier Studien erfüllten alle acht in der Validitätsprüfung abgefragten Qualitätskriterien (eindeutige und transparente Beschreibung der Studienpopulation; valide und reliable Erfassung der Exposition (QUS!); valide und reliable Erfassung der patientennahen Zielgröße (Fraktur); ausreichend lange und vollständige Nachbeobachtung und bei der Analyse der Ergebnisse, die Adjustierung für bekannte wichtige prognostische Faktoren (sh. auch SOP 3 und Stufenpläne im Anhang). Lediglich bei einer Publikation (Huang et al. 1998) musste auf die Möglichkeit, die Arbeit trotz eines methodischen Mangels einzuschließen, zurückgegriffen werden: in der Publikation wird die Follow-Up Rate nicht explizit berichtet, da es sich um eine Subgruppenuntersuchung einer andauernden prospektiven populationsbezogenen Studie handelt (Hawaiian Osteoporosis Study) (Heilbrunn et al. 1985). Da diese Daten die einzigen mit Populationsbezug waren, haben wir sie dennoch berücksichtigt. Zwei prospektive Studien wurden aus methodischen Gründen aus der Analyse ausgeschlossen [bei einer Studie stellte sich auf den zweiten Blick heraus, dass es sich um ein "nested case-control" Design handelte (Garnero et al. 1998); die zweite stellte die prospektiven Daten unvollständig lediglich im Diskussionsteil der Arbeit dar (Dretakis et al. 1998)].

Auch unter den eingeschlossenen Studien war noch erhebliche epidemiologische, klinische und technische Heterogenität festzustellen:

Die Studienpopulationen waren so heterogen, dass ein direkter Vergleich in nur sehr eingeschränktem Umfang möglich war. Zwei Studien nutzten sekundäranalytisch Daten großer prospektiver Studien (Dargent-Molina et al.: EPIDOS und Huang et al.: HOS). Bis auf die Publikation von Huang et al. (1998) (s.o.), berichten alle Arbeiten Daten aus Hochrisikopopulationen. In den Studien von Pluijm et al. (1999) und Dargent-Molina et al. (1999) sind die Populationen aufgrund ihres hohen Alters (im Mittel 82.8 Jahre bzw. 80.5 Jahre) als hochgradig frakturgefährdet anzusehen. Darüber hinaus handelte es sich in der Studie von Pluijm um eine gemischt-geschlechtliche Population, in die Studie von Dargent-Molina (EPIDOS-Studie) wurden nur Frauen eingeschlossen. In der Arbeit von Stewart et al. (1999) ergab sich die hochgradige Frakturgefährdung aus der Anamnese einer vorangegangenen Hüftfraktur. Die Studienpopulation (postmenopausale Frauen) von Gnudi et al. (2000) enthielt einen relevanten Anteil (ca. 50 %) von bereits behandelten Patientinnen mit Osteoporose.

Auch bei der Expositionsmessung, die zwar in allen fünf Studien transparent dokumentiert wurde und intern den Ansprüchen an Reproduzierbarkeit und Validität genügte, war erhebliche Heterogenität festzustellen. Vier Verfahren unterschiedlicher Gerätehersteller, die sich in Messort (Kalkaneus, Radius, Patella), Schalleinleitung (Wasser, Gel) und Messgrößen (SOS, BUA, UTV) unterschieden, liefern nicht direkt vergleichbare Werte. Eines der eingesetzten Verfahren ist bereits technisch überholt und obsolet (Gnudi et al. 2000).

In vier Studien war die Hüftfraktur primäre Zielgröße, in zwei Studien sogar einzige Zielgröße (Dargent-Molina et al. 1999 und Stewart et al. 1999). Die Erfassung erfolgte in postalischen oder telefonischen Befragungen, die Patientenangaben wurden im Falle einer positiven Antwort durch Einsichtnahme in medizinischen Unterlagen verifiziert. Lediglich die populationsbezogene Untersuchung von Huang et al. (1998) nimmt keine Auswertung von Hüftfrakturereignissen vor, sondern analysiert die Daten nach Wirbel- und nicht-spinalen Frakturen.

Auch die berichteten Effektgrößen unterschieden sich in den Studien erheblich. In einer Publikation (Dargent-Molina 1999) werden Angaben zur Sensitivität des QUS-Verfahrens zur Erkennung von Frakturgefährdung gemacht. Da es sich um prospektive Daten handelt, können diese Größen als direkte Aussage zum prognostischen Wert interpretiert werden. Vier Studien gaben adjustierte relative Risiken an. Dabei wurden in einer Studie nur das Alter, einer weiteren Studie zusätzlich Gewicht und Geschlecht und in der letzten Studie Alter, Gewicht, Geschlecht, Körpergröße und laufende Osteoporosetherapie berücksichtigt. Die Arbeit von Huang et al. (1998) gab altersadjustierte Odds ratios an. Rohdaten, die eine Nachberechnung von einheitlichen Effektgrößen zulassen, sind in den Publikationen nicht angegeben; diese für meta-analytische Zwecke von den Autoren zu erbitten, übersteigt den Rahmen eines Kurz-HTA.

Vor diesem heterogenen Hintergrund waren weitergehende meta-analytische Auswertungen nicht möglich (vergl. auch Altman 2001).

Damit scheint uns die Basis der folgenden Diskussion und Schlussfolgerungen vergleichsweise schmal. Unsere hierarchisch strukturierten Recherchen (vergl. SOP 2) fanden für den Zeitraum September 1998 (nach Redaktionsschluss für den HTA-Bericht von Homik & Hailey) bis Februar 2001 lediglich die fünf oben präsentierten prospektiven Studien zu unserer Kernfragestellung.

Dennoch lassen sich aus dieser begrenzten Evidenzlage durchaus relevante Schlussfolgerungen ableiten. Sie beziehen sich v. a. auf die Formulierung von Forschungsbedarf und auf den Verbraucher- und Patientenschutz. Daneben werden die Aussagen des vorliegenden HTA-Berichtes gestützt.

II C.6 Diskussion der Forschungsfragen

1. Wie valide lässt sich das Frakturrisiko mithilfe der quantitativen Ultraschall-Verfahren im Vergleich zum DXA-Verfahren abschätzen?

Osteodensitometrische Verfahren sind geeignet einen Beitrag zur Diagnostik, Frakturrisikoabschätzung und Verlaufskontrolle bei Osteoporose zu liefern. Die Rolle der QUS in diesem Kontext lässt sich noch nicht abschließend beurteilen.

Die Osteoporose ist eine systemische Skeletterkrankung, die durch niedrige Knochenmasse und eine Verschlechterung der Mikroarchitektur des Knochengewebes mit daraus folgender Zunahme der Knochenbrüchigkeit und gesteigertem Frakturrisiko charakterisiert ist.

Die Diagnose der Osteoporose wird heute anhand der Knochendichte (gemessen mit dem DXA-Verfahren) gestellt. Dabei werden Werte zwischen 1 und 2.5 Standardabweichungen unterhalb des Mittelwertes einer jungen gesunden Referenzpopulation als "osteopenisch", mehr als 2.5 Standardabweichungen unterhalb des Mittelwertes als "osteoporotisch" und - wenn zusätzlich Frakturen vorliegen als "manifest osteoporotisch" bezeichnet. Dieses Hilfskonstrukt wird herangezogen, weil die Knochendichte zur Zeit der einzige Einflussfaktor ist, der einigermaßen valide, präzise und nicht-invasiv gemessen werden kann. Für eine korrekte Diagnose der Frakturanfälligkeit wäre allerdings eine genaue Messmöglichkeit von Knochenfestigkeit (definiert über Knochenmaterialeigenschaft, Knochendichte, Geometrie, Mikrofrakturen, Anzahl, Tiefe und Ausdehnung der Resorptionslakunen) wichtig, da die Verminderung bzw. Verschlechterung dieser Faktoren die Ursache für eine verringerte Bruchfestigkeit darstellt. Darüberhinaus spielen selbstverständlich auch das Sturzrisiko und die durch einen Sturz entstehende Kraffteinwirkung eine entscheidende Rolle bei der Frakturentstehung.

Die Bruchfestigkeit kann jedoch (noch) nicht in vivo bestimmt werden: Man ist also auf intermediäre oder Surrogatparameter angewiesen. Zu diesen zählen u.a. konkurrent erhobene Merkmale der Knochenqualität wie Kalksalzgehalt, biochemische Qualität der Grundsubstanz, radiologisch dargestellte Mikroarchitektur oder Knochenhistologie. Ein weiterer Indikator der zur Abschätzung der Bruchfestigkeit (und als Kenngröße der prognostischen Validität von Osteodensitometrieverfahren) herangezogen wird ist die inzidente oder prävalente Fraktur. Aber auch diese ist ein Surrogatmerkmal, da sie nicht nur von der Knochenqualität, sondern zusätzlich noch von der Art und Stärke eines möglichen Traumas abhängt.

Über die Jahre ist eine Vielzahl unterschiedlicher osteodensitometrischer Verfahren entwickelt worden (Genant et al. 1998). Am weitesten verbreitet ist die Zwei-Spektren-Röntgenabsorptiometrie (engl. Dual X-Ray Absorptiometry, DXA), die die "Knochendichtebestimmung" an der Lendenwirbelsäule und am proximalen Femur, d. h. an relevanten Lokalisationen osteoporosebedingter Frakturen, ermöglicht (Blake et al. 1999). Ein Nachteil der Methode liegt darin, dass die Dichtemessung nicht di-

rekt erfolgen kann sondern aus Projektionsfläche und Flächenbelegung berechnet wird. Weiter stören degenerative Veränderungen der LWS (Spondylosis deformans), Osteochondrose (Bandscheibendegenerationen), Skoliose und Torsion, sowie Verkalkungen der Aorta und bereits eingetretene Wirbelkörperfrakturen die Messung erheblich. Daher wird bei älteren Menschen (> 65 J.) wegen der zunehmenden Häufigkeit der Spondylophyten die DXA-Untersuchung der Hüfte empfohlen.

Als von ionisierenden Strahlen freie Technik zur Knochendiagnostik wurde im Laufe der letzten Jahre die quantitative Ultraschalluntersuchung (QUS) entwickelt (Glüer et al. 1998). Ultraschallmessungen werden ausschließlich an peripheren Skelettstrukturen wie dem Kalkaneus, der Patella, der Tibia oder an einzelnen Fingerknochen durchgeführt. In der berücksichtigten Literatur stehen Messungen am Kalkaneus ganz im Vordergrund. Im Gegensatz zur Wirbelsäule und dem Schenkelhals sind diese Messregionen frei zugänglich, und ein direkter Kontakt durch die Haut ist möglich. Der Nachteil bei dieser Methode besteht darin, dass die bei den Messungen herangezogenen Skelettareale nicht prädisponiert sind für osteoporosespezifische Frakturen.

Ergebnisse von QUS- und DXA-Messungen sind in aller Regel positiv korreliert; die Korrelationen erreichen in-vivo $r = 0.80$, in in-vitro-Studien liegen sie auch darüber. Klinische Studien berichten zumeist deutliche niedrigere Korrelationswerte (Bereich 0.30 – 0.50) wobei nicht immer klar wird, ob die Messungen am gleichen Messort durchgeführt wurden (Homik & Hailey 1998). Es ist davon auszugehen, dass Ultraschallverfahren nur zum Teil die gleichen Knocheneigenschaften abbilden wie die radiologischen Densitometrieverfahren (Barkmann et al. 1999).

Auf der Basis eines HTA-Berichtes aus 1998 und fünf ergänzender Originalarbeiten aus den Jahren 1998 bis 2000 ergibt sich folgende Antwort auf die eingangs gestellte Frage:

Prospektive Kohortenstudien zur Vorhersage von Wirbel- und Schenkelhalsfrakturen wurden für die QUS-Messungen mit unterschiedlichen Techniken an unterschiedlichen Messorten (oft am Kalkaneus) durchgeführt. Eine prognostische Gleichwertigkeit der Ultrasonometrie gegenüber DXA wurde belegt, QUS und DXA ermitteln, jeweils für sich betrachtet, vergleichbare relative Risiken für Frakturen innerhalb eines Beobachtungszeitraums von $\frac{1}{2}$ bis $5 \frac{1}{2}$ Jahren. Die Meta-Analyse von Marshall et al. (1996) berichtet für Knochendichtewerte pro Erniedrigung um je 1 Standardabweichung ein relatives Risiko von 1.6 für alle Frakturen und alle Messorte. Der in den fünf neueren prospektiven Studien berichtete Bereich der relativen Risiken (bezogen auf eine Differenz von - 1 SD) Abnahme liegt für die BUA-Messungen in der gleichen Größenordnung (von 1.17 bis 2.3). Knochendichte und Ultraschallparameter scheinen von einander unabhängige Risiken zu bezeichnen (Homik & Hailey 1998). Die Kombination von QUS und DXA bei multivariater Betrachtung führt möglicherweise zu einer verbesserten Frakturprädiktion, die uns vorliegende Datenbasis zu dieser Frage ist allerdings begrenzt. In der Studie von Gnudi et al. (2000) steigt bei Kombination der QUS und BMD Parameter in multivariaten

Berechnungen das relative Frakturrisiko um das Fünffache. Die Unsicherheit dieser Modellberechnungen ist an den extrem weiten Konfidenzintervallen abzulesen (RR 13.46 (95 % KI 2.25 - 80.53) für BMD-DR + UTV-DR; RR 16.93 (95 % KI 2.22 - 128.84) für BMD-DR + UTV-P). In der Arbeit werden heute obsoleete Verfahren sowohl zur QUS- als auch zur Knochendichtemessung verwendet. Hans et al. (1997) weisen in einer Übersichtspublikation zu EPIDOS-Ergebnissen ebenfalls auf die Möglichkeit hin, durch Kombination der beiden Verfahren eine Hochrisikogruppe zu identifizieren, ihre Schätzungen lagen aber lange nicht so hoch (ca. 1.8 - fach). Empfehlungen, welche Verfahren und Messorte zur Prädiktion welcher Frakturtypen herangezogen werden könnten, lassen sich aus diesen Daten allerdings noch nicht ableiten, hier ist sicherlich weiterer Forschungsbedarf.

Eine prognostisch valide Früherkennung ist dann besonders sinnvoll, wenn sie zu einer effektiven Frühtherapie führt. Das Problem der QUS-Messungen in diesem Zusammenhang ist, dass die durch sie erfasste Knocheneigenschaft unklar ist. Dies macht die Entwicklung spezifischer Therapieverfahren nahezu unmöglich. Es war uns in diesem Bericht nicht möglich zu prüfen, ob und wie weit bekannte Therapieverfahren zu einer Veränderung der mit QUS gemessenen Parameter führen.

2. Welches der unterschiedlichen QUS-Verfahren ist zur Vorhersage von Frakturen am besten geeignet?

Diese Eignung ist abhängig von der Zuverlässigkeit der Messungen und - in Abhängigkeit vom Messergebnis - der Höhe und Sicherheit des geschätzten Relativen Risikos für Frakturen. Aus praktischer Sicht sind schließlich die Verfügbarkeit und Verbreitung der Technik zu berücksichtigen. Das in der Studie von Gnudi et al. (2000) verwendete Gerät Signet Device (Osteotechnology Inc. Framingham) ist auf dem Markt nicht mehr verfügbar.

Im vergangenen Jahrzehnt ist eine Vielzahl unterschiedlicher quantitativer QUS-Verfahren entwickelt worden (Njeh et al. 1999). "Diese sollten hinsichtlich ihrer Eignung und Validierung nicht über einen Kamm geschoren werden..." (Glüer und The International Quantitative Ultrasound Consensus Group 1997).

Die meisten zur Zeit verfügbaren Ultraschallgeräte messen die Transmissionsgeschwindigkeit des Schalls durch den Knochen. Bei der Transmission werden nicht nur die Geschwindigkeit, sondern auch die Abschwächung gemessen. Die quantitativen Parameter sind die Ultraschallgeschwindigkeit (SOS) und die Ultraschallabschwächung (BUA). Einige Geräte verwenden Kombinationsparameter, die z. B. als „Steifigkeit“ angegeben werden. Diese sollten nicht mit den biomechanischen Parametern verwechselt werden. Die meisten Geräte messen am Kalkaneus: Achilles von der Firma Lunar, UBA 575 äquivalent Sahara der Firma Hologic und CUBA von der Firma McCue Ultrasonic.

Neuere Geräte (z. B. UBIS 3000 von DMS) haben die Möglichkeit, die Messregion bildlich darzustellen und versuchen, dadurch die Präzision zu erhöhen. Ob dies für die klinische Anwendung ein Vorteil ist, ist nicht gesichert, außerdem fehlen ausrei-

chend große Referenzkollektive. Als weiterer Messort fungiert die distale Metaphyse der proximalen Phalangen der langen Finger II-V, verwendet wird das Gerät DBM Sonic 1200 der Firma IGEA. Vorwiegend kortikaler Knochen wird erfaßt bei der Messung der Ultraschallreflexionsgeschwindigkeit an der Tibia mittels Soundscan 2000 von Myriad Ultrasound Ltd.. Die gleiche Messgröße wird an unterschiedlichen Knochen mit dem Gerät Omnisense von Sunlight gemessen.

Der gebräuchlichste Messort ist der Kalkaneus, vermessen in transversaler Transmission. Ultraschallmessungen an den Fingerphalangen, ebenfalls durchgeführt in transversaler Transmissionstechnik, sind das am zweitbesten untersuchte Verfahren im Ultraschallbereich (Wüster et al. 2000), das allerdings weder in unserem HTA-Bericht noch in den untersuchten fünf Studien berücksichtigt wurde.

Diese Aufzählung verdeutlicht das gravierende Problem der Heterogenität der Techniken, Messgrößen und Ergebnisse und damit deren eingeschränkte Vergleichbarkeit. Hinzu kommen heterogene Referenzkollektive und –bereiche. Eine übergreifende Standardisierung ist bisher, anders als bei DXA, nicht erfolgt.

Sie verdeutlicht auch die Schwierigkeit einer eindeutigen Antwort auf die gestellte Frage: Berücksichtigt man Reliabilität, prognostischen Wert und Verbreitung gleichzeitig, dann spricht vieles für die transversale Messung von BUA/SOS am Kalkaneus.

3. Sollte die Analyse ergeben, dass die quantitative Ultraschalldiagnostik eine valide und zuverlässige Abschätzung des Frakturrisikos erlaubt: welche qualitätssichernden Maßnahmen sind dann vor einem Einsatz in der ärztlichen Praxis zu realisieren, um zu reliablen und unverzerrten Ergebnissen zu kommen ?

Eine wichtige Voraussetzung für die Einführung eines Ultraschall-Densitometriegerätes in Deutschland in die klinische Praxis sollte das Vorhandensein eines ausreichend großen (Power Berechnung!) repräsentativen Referenzkollektivs junger gesunder Frauen und Männer aus unserer europäischen Region sein. Dabei ist zu beachten, dass Referenzwerte für alle einzuführenden Systeme und für jede Messlokalisierung getrennt zu erheben sind. Zum Beispiel umfassen die Referenzpopulationen für Frauen für den Achilles ca. 5.000, für den Soundscan 2000 ca. 5.000 und für den DBM Sonic 1200 ca. 10.000 Frauen (Wüster 1999). Die männlichen Referenzkollektive sind bei allen Geräten wesentlich kleiner. Bei kleineren Kollektiven ist das Problem der Generalisierbarkeit auf lokale Populationen zu berücksichtigen. In der EVOS-Studie wurde immerhin gezeigt, dass mittlere Knochendichtewerte an der LWS, gemessen mit DXA, großen geographischen Schwankungen zwischen Nord- und Südeuropa und auch dem Osten und Westen Europas unterworfen sind. Praktisch sollte jeder Messende auch gesunde Personen messen und mit dem Referenzkollektiv seines Gerätes überprüfen, bevor er Rückschlüsse aus der Messung für eigene Patienten zieht.

Bei DXA-Messungen hat sich international der Gebrauch von T- und Z-Werten eingebürgert, um zu kategorialen Urteilen (z. B. „Osteopenie“, „Osteoporose“ nach

WHO) zu kommen. Hiervon scheinen die QUS-Messungen angesichts der geschilderten Heterogenitäten noch weit entfernt. So wird man in diesem Gebiet vorsichtig mit kategorialen Einteilungen sein müssen, auch wenn in der Praxis gerne mit „grünen“, „gelben“ und „roten“ Messwertbereichen hantiert wird. Bereits kleinste Veränderungen sowohl des Mittelwertes als auch der Standardabweichung (SD) eines Referenzkollektives bringen große Veränderungen im Prozentsatz der "Erkrankten" mit sich. In keinem Falle wäre es heute schon gerechtfertigt, bei Messwerten unterhalb von -2.5 SD eines Messwerts von einer „Osteoporose“ zu sprechen.

Solange der Gegenstand der Messung unklar bleibt, wird es auch schwer sein, ein Phantom (European Spine Phantom bei DXA; WHO 1994) zur Kreuzkalibrierung und fortwährenden Qualitätssicherung auszuarbeiten.

Als Qualitätssicherungsmaßnahmen zum korrekten Einsatz der QUS-Verfahren im Rahmen einer umfassenden Osteoporosediagnostik werden heute in Deutschland vom Dachverband der wissenschaftlich-osteologischen Fachgesellschaften folgende Grundvoraussetzungen genannt (nach: Osteoporose 2000, Unterlagen zum Zertifizierungskurs):

- Einsatz und Interpretation von QUS nicht isoliert, sondern nur im Kontext von Anamnese und klinischem Befund (körperliche Untersuchung, Röntgenbild von BWS und LWS, Labor Initialprogramm). Im Einzelfall kann eine weitere apparative Untersuchung (DXA) zur Feststellung eines Hochrisikostatus sinnvoll sein. (Kommentar: diese Forderung sollte nicht regelhaft zu einem additiven Gebrauch mehrerer Techniken, sondern möglichst zu ihrer jeweiligen Substitution führen.)
- Einsatz von validierten Geräten
- Messung durch zertifiziert geschultes Personal
- Interpretation durch zertifiziert geschulten Arzt
- Arbeitstäglige Phantommessung (mit knochenäquivalentem Prüfkörper) und dokumentierte Analyse der Ergebnisse.

4. Welchen weiteren Forschungsbedarf zur Klärung des Stellenwertes der Ultraschalldiagnostik im Rahmen der Frakturprävention gibt es?

Die DFG (Deutsche Forschungsgemeinschaft) unterscheidet in ihrer Denkschrift zur Klinischen Forschung (2000) drei verschiedene Typen, die grundlagenorientierte, die krankheits- und die patientenorientierte klinische Forschung.

Angesichts der vielfältigen Unklarheiten und Unsicherheiten im Bereich der QUS-Verfahren scheint eine Fortführung, eventuell auch Intensivierung der medizinischen und physikalischen Grundlagen- und krankheitsbezogenen Forschung erforderlich (Wüster 1999). Dies würde auch helfen, das diagnostische Potential von QUS zu klären.

Im Anschluss halten wir die Durchführung von Interventionsstudien für unverzichtbar. Bislang liegen nur valide Daten zur Vorhersagbarkeit von Frakturen bei Personen vor, die bislang **nicht** mit knochenwirksamen Substanzen behandelt wurden. Prinzipiell ist festzuhalten, dass eine diagnostische Maßnahme nur dann sinnvoll ist, wenn zumindest die Option einer aus dem Messergebnis sich ableitenden – effektiven und effizienten – Therapie besteht. Solche Optionen können sich auf die Primär-, Sekundär- oder Tertiärprävention der osteoporoseassoziierten Fraktur beziehen. Über keinen dieser Ansätze besteht bisher eine ausreichende Klarheit (obwohl diese Frage nicht Gegenstand dieses Kurz-HTA-Verfahrens war). Diese Interventionsstudien müssten auch zeigen, dass die heute zur Verfügung stehenden Therapeutika (Hormonersatztherapie, SERM, Vitamin D und Calcium, Bisphosphonate, Fluor u.a.) QUS-Parameter messbar und klinisch relevant beeinflussen.

Erst dann liesse sich beurteilen, ob sich QUS auch zur Therapiekontrolle und Verlaufsbeobachtung des natürlichen oder klinischen Verlaufs der Osteoporose und ihrer Risiken eignet. Sollte dies zu bejahen sein, dann könnte es auch der Förderung der Patientencompliance zugute kommen. In uns bekannten Behandlungstudien wurde bisher kein Effekt von Monitoring mittels Knochendensitometrie und Knochenstoffwechsellmarkern auf die Therapietreue berichtet, hier muß weiter geforscht werden.

Aus der Sicht von Glüer et al. (2000) sollten vor einem Einsatz der QUS-Verfahren weitere Validierungsstudien durchgeführt werden: In vitro-Untersuchungen und theoretische Modelle, um besser zu verstehen, welche Charakteristika des Knochens von den neuen Verfahren erfasst werden; Querschnittsstudien, in denen die neuen Verfahren parallel zu den Standardverfahren eingesetzt werden, und letztlich prospektive Langzeitstudien, in denen die Vorhersagekraft der Verfahren bezüglich des Frakturrisikos bestätigt werden kann.

Der Konsensusbericht der amerikanischen National Institutes of Health (2001) nennt eine ganze Reihe von Fragestellungen, die er in der zukünftigen Forschung für vorrangig hält. Aus der langen Liste sei nur der Forschungsbedarf bezüglich der QUS-Verfahren im Rahmen der Frakturprävention erwähnt:

„.....mehr müßten wir hinsichtlich des Frakturrisikos über das Zusammenspiel von knochenabhängigen und knochenunabhängigen Faktoren wissen. Ein verbindlicher Algorithmus müsste entwickelt werden unter Einbezug von Risikofaktoren und Knochendichtemessungen und Surrogatmarkern für den Knochenumbau.....“. Gegenwärtige Forschungen zielen darauf hin, ein komplexes Modell zu entwickeln, in dem die verschiedenen Risikoindikatoren zusammengefasst werden und aus denen dann ein absolutes Frakturrisiko über die nächsten 5 oder 10 Jahre abgeleitet werden kann (Kanis et al. 2000).

5. Wie wären die unter 1 – 3 ermittelten Ergebnisse zur Patienteninformation und zum Verbraucherschutz zu nutzen?

Patienten bzw. Verbraucher können eine informierte Entscheidung nur dann treffen, wenn ihnen als Grundlage eine evidenzbasierte und für sie verständliche Information zur Verfügung steht. Dies bedeutet, dass die Informationen dem aktuellen Forschungsstand entsprechen und in geeigneter Weise vermittelt werden müssen. Der Wunsch des Patienten bzw. Verbrauchers nach Informationen und seine Fähigkeit, diese zu nutzen, werden oft unterschätzt, und sowohl Risiken als auch Unsicherheiten medizinischer Verfahren werden nicht kommuniziert (v. Stackelberg et al. 2001).

Informationen zu QUS erreichen die Verbraucher und Patienten aktuell auf vielen Wegen (z. B. Zeitungsannoncen, Internet, Aushänge in Wartezimmern etc.). Sie zielen darauf ab QUS als IGEL-Leistungen anzubieten. Die Osteosonometrie ist bisher nicht Bestandteil des EBM und muss von den Nachfragenden privat finanziert werden.

Einige in diesem Zusammenhang typische Auskünfte lesen sich folgendermaßen:

-international anerkannte Methode zur Beurteilung des Frakturrisikos im Rahmen der Osteoporosediagnostik; Diagnose Osteoporose ist anhand der Knochendichtemessung per QUS zu stellen; Ultraschalluntersuchung erlaubt eine noch frühere Erkennung des Verlustes an Knochendicke und –elastizität als herkömmliche Verfahren; Messverfahren mit Röntgenstrahlen oder Ultraschallsignalen an den verschiedenen Messstellen sind gleichwertige Untersuchungstechniken; Ultraschallknochendichtemessung stellt einen wesentlichen Beitrag zur Früherkennung einer Osteoporose und zur Einleitung einer gezielten Therapie dar; durch die Messung lässt sich der Erfolg einer eingeleiteten Therapie gut demonstrieren; jährliche Verlaufskontrollen bei Frauen über 40 erforderlich.....

Dagegen müsste eine seriöse Information folgende Sachverhalte festhalten:

-Verfahren noch in der Entwicklung; hervorragende Messeigenschaften unter optimierten Laborbedingungen; unter klinischen Alltagsbedingungen noch viele Fragen offen: Gemessenes unklar; Qualitätssicherung im Fluß; Verlaufskontrolle unsicher; therapeutische Konsequenzen offen; schützt nicht gegen Fraktur; nicht jeder mit schlechten Werten erleidet Fraktur.. (gilt auch für DXA), prognostischer Wert, Interaktion mit DXA.....

Um eine Einschätzung des individuellen Frakturrisikos zu ermöglichen, sind Angaben zum altersabhängigen Risiko an Osteoporose zu erkranken erforderlich. Wirksamkeit, Risiken und Folgen der quantitativen Ultraschalluntersuchung des Knochens müssen durch die Verbraucher beurteilt werden können.

Hier eröffnet sich dem Patienten- und Verbraucherschutz ein lohnendes Feld.

Ein Schritt in Richtung valide Patienteninformation stellt sicherlich die DISCERN-Initiative dar. Im Rahmen dieses Projektes wurden Kriterien erarbeitet (DISCERN-

Kriterien), die sicher stellen sollen, dass Verbraucher-Informationsbroschüren einem methodisch-qualitativen Standard entsprechen. Die Original-Version wurde 1997 von der British Library und der University of Oxford erstellt, die deutsche Fassung wurde 2000 von der Abteilung Epidemiologie, Sozialmedizin und Gesundheitssystemforschung der Medizinischen Hochschule Hannover und von der Ärztlichen Zentralstelle für Qualitätssicherung herausgegeben (www.discern.de/instrument.htm).

DISCERN besteht aus 15 Schlüsselfragen sowie einer Bewertung der Gesamtqualität einer Patienteninformationsschrift. Jede der 15 Schlüsselfragen repräsentiert ein eigenes Qualitätskriterium - d. h. ein entscheidendes Merkmal qualitativ guter Informationen über Behandlungsalternativen.

Eine qualitativ gute Publikation über Behandlungsalternativen wird:

1. Klare Ziele haben
2. Ihre Ziele erreichen
3. Für den Nutzer bedeutsam sein
4. Ihre Informationsquellen klar benennen
5. Das Erstellungsdatum der Informationen klar benennen
6. Ausgewogen und unbeeinflusst sein
7. Zusätzliche Informationsquellen aufführen
8. Auf Bereiche von Unsicherheit hinweisen
9. Die Wirkungsweise eines Behandlungsverfahrens beschreiben
10. Den Nutzen eines Behandlungsverfahrens beschreiben
11. Die Risiken eines Behandlungsverfahrens beschreiben
12. Die Folgen einer Nicht-Behandlung beschreiben
13. Die Auswirkungen von Behandlungsverfahren auf die Lebensqualität beschreiben
14. Verdeutlichen, dass mehr als ein mögliches Behandlungsverfahren existieren könnte
15. Eine partnerschaftliche Entscheidungsfindung (shared decision-making) unterstützen

Zum jetzigen Zeitpunkt muß man jedoch konstatieren, dass die derzeit im Umlauf befindlichen Informationsmaterialien den "DISCERN" Kriterien nur in einem sehr eingeschränkten Maße entsprechen. Eine adäquate Kosten-Nutzen-Abwägung (für ihren Einzelfall) ist den Betroffenen somit kaum möglich.

II C.7 Schlussfolgerungen

Die quantitative Ultraschalluntersuchung des Knochens ist eine relativ neue, attraktive Methode, deren sinnvoller Einsatz sich derzeit anhand der vorhandenen wissenschaftlichen Daten noch nicht abschließend beurteilen lässt. Sie ist bisher nicht Bestandteil des EBM, erfährt aber doch zunehmende Verbreitung.

Mit Ultraschall wird nicht, wie bei DXA, absorptiometrisch der Kalksalzgehalt des Knochens (als Indikator der Bruchfestigkeit) gemessen, sondern in ihrer biologischen Referenz noch unklare Parameter wie die Abschwächung und Geschwindigkeit des in Knochen eingebrachten Ultraschalls. Die Messparameter SOS und BUA korrelieren moderat bis gut (in Abhängigkeit von der Messlokalisation und der Versuchsanordnung) mit Knochendichtewerten. Mittels des quantitativen Ultraschallverfahrens wird vermutlich ein anderes Charakteristikum des Knochens gemessen als mit der Dichtemessung.

Die Eignung dieses Verfahrens ist abschließend noch einmal unter den folgenden drei Gesichtspunkten zu analysieren:

- Beitrag zur Diagnose der Osteoporose
- Abschätzung des Frakturrisikos
- Verlaufs- und Therapiekontrolle

1. Einsatz zur Diagnose

Die Diagnose einer Osteoporose wird klinisch in Verbindung mit einer Knochendichtemessung, einem Röntgenbild und laborchemischen Basisdaten gestellt. Zur Klassifikation von Kranken wird heute, auch aus praktischen Gründen, auf die Ergebnisse der Knochendichtemessungen zurückgegriffen (WHO 1994). Sie erlauben eine Abschätzung des Kalksalzgehalts des Knochen und damit einer Determinante seiner Bruchfestigkeit, die wiederum nur ein Prädiktor der Frakturzinzidenz ist.

Solange der Gegenstand der verschiedenen QUS-Messungen nicht klar ist, kann die Frage nach dem Beitrag des Verfahrens zur Osteoporosediagnose nicht beantwortet werden. Zum jetzigen Zeitpunkt ist jedenfalls die Gleichsetzung von erniedrigten BUA- und SOS- etc. Werten und Osteoporose nicht zu halten. Hier besteht dringend Bedarf zu weiterer medizinischer und physikalischer Forschung.

Es besteht zum jetzigen Zeitpunkt keine Indikation, Ultraschalluntersuchungstechniken in der Osteoporose-Diagnostik einzusetzen. Hierüber müssen Verbraucher und Patienten aufgeklärt werden.

2. *Frakturrisikoabschätzung*

Die Eignung von QUS-Verfahren zur prospektiven Abschätzung des Frakturrisikos wurde in einer Reihe von prospektiven Studien bestätigt. Die meisten Untersuchungen beziehen sich hierbei auf die Ultraschallmessungen am Kalkaneus, es ist das am besten untersuchte und validierte Verfahren. Der prognostische Wert von QUS-Verfahren ist ähnlich dem von DXA-Verfahren. Auch wenn QUS-Parameter unabhängige und zusätzliche prognostische Informationen zu geben scheinen, ist es bis heute nicht sicher, ob QUS DXA-Messungen ersetzen kann oder ob sie als additive Technologie bewertet werden muss und welchen Zusatznutzen sie in diesem zweiten Falle hätte.

Ein Einsatz von QUS-Verfahren für die Frakturrisikoabschätzung scheint möglich. Ihr Stellenwert für die Risikoerkennung und klinische Prävention ist letztlich noch unsicher. Ihr Gebrauch sollte vorzugsweise an wissenschaftliche Fragestellungen und vor allem auch an eine konsequente Qualitätssicherung gebunden sein. Über Abweichungen hiervon sind diejenigen zu informieren, die das Verfahren, etwa als IGeL-Leistung, in Anspruch nehmen wollen.

3. *Verlaufs- und Therapiekontrolle*

Die Datenlage für einen Einsatz der QUS-Verfahren in der Verlaufskontrolle scheint begrenzt und kontrovers. Sichere Aussagen sind zur Zeit nicht möglich, auch weil die therapeutische Beeinflussung und Veränderungssensitivität der QUS-Parameter zu wenig geklärt sind. Einschränkend ist zu betonen, dass dieser Teilaspekt nicht Gegenstand unseres HTA-Schnellverfahrens war.

Die jetzige Datenlage berechtigt nicht, Empfehlungen zur periodischen Messung von QUS-Parametern (etwa in jährlichem Abstand) abzugeben.

Aktualisierungsbedarf

Eine Aktualisierung des Berichtes wird erst erforderlich, wenn sich die Evidenzlage deutlich geändert hat. Dies zu beurteilen, setzt ein formalisiertes Screeningverfahren in höchstens einjährigem Abstand voraus. Hier sehen wir eine Aufgabe von DIMDI, die bisher möglicherweise nicht ausreichend reflektiert worden ist.

II D Literatur

1. Alenfeld FE. Ultrasound measurements at the proximal phalanges in healthy women and patients with hip fractures. *Osteoporos Int* 8: 693-698;1998
2. Altmann D, in: *Systematic Reviews in Health Care. Meta-analysis in context.* ed. by: Egger M, Smith G, Altmann DG. *BMJ Sec.edit.*: 228-247; 2001
3. Baran DT, Kelly AM, Karellas A, Gionet M, Price M, Leahey D. Ultrasound attenuation of the os calcis in women with osteoporosis and hip fractures. *Calcif Tissue Int* 43: 138-42; 1988
4. Barkmann R, Kantorovich E, Singal C. A new method for quantitative ultrasound measurements at multiple skeletal sites – first results of precision and fracture discrimination. *J Clin Densitom*; 1999
5. Barkmann R, Heller M, Glüer CC. Methoden der In Vivo-Ultraschall-Messtechnik am Skelett: Grundlagen und Technische Realisierung. *J Miner Stoffwechs* 4: 22-27; 1999
6. Barkmann R. *Fortbildungshandbuch Osteoporose 2000*, Berlin, 2000
7. Bauer DC, Glüer CC, Genant HK, Stone K. Quantitative ultrasound and vertebral fracture in postmenopausal women. *J Bone Miner Res* 10: 353-8; 1995
8. Bauer DC, Glüer CC, Cauley JA, for the Study of Osteoporotic Fractures Research Group: Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women. A prospective study. *Arch Intern Med* 157: 629-34;1997
9. Blake GM, Glüer CC, Fogelman I. Bone Densitometry: current status and future prospects. *Br J Radiol* 70 : S177-S186; 1997
10. Blake JM, Ryan ED, Beaumont LF, Webber CE. Changes in spine and radius bone density during long term hormone replacement. *Can J Physiol Pharmacol* 77: 505-509; 1999
11. Bouxsein ML, Courtney AC, Hayes WC. Ultrasound and densitometry of the calcaneus correlate with the failure loads of cadaveric femurs. *Calcif Tissue Int* 56(2): 99-103; 1995
12. Chappard C. Assessment of the relationship between broadband ultrasound attenuation and bone mineral density at the calcaneus using BUA imaging and DXA. *Osteoporos Int* 7: 316-322; 1997
13. Coester A, Haberkamp M, Allolio B. Inzidenz von Schenkelhalsfrakturen in der Bundesrepublik Deutschland im internationalen Vergleich. *Soz Präventivmed* 39: 287-292; 1994
14. Consensus Development Conference: Diagnosis, prophylaxis and treatment for Osteoporosis. *American Journal of Medicine* 90: 107-110; 1991
15. Cooper C, Atkinson E, o'Fallon M, Melton III LJ. Incidence of Clinically diagnosed vertebral fractures: A population based Study in Rochester, Minnesota, 1985-89. *J Bone Miner Res* 7(2): 221-227; 1992

16. Cooper C, Atkinson EJ, Jacobsen SJ, o'Fallon MW, Melton III LJ: Population Based Study of Survival after Osteoporotic Fractures. *Am J Epidemiol* 137(9): 1001-1005; 1993
17. Cooper C. Epidemiology of vertebral fractures in western populations. *Spine* 8(1): 1-11;1994
18. Cummings SR, Nevitt MC, Browner WS. Risk factor for hip fracture in white women. Study of Osteoporotic Fractures Research Group. *N Engl J Med* 332: 767-773; 1995
19. Dargent-Molina P, Schott AM, Hand D, Favier F, Grandjean H, Baudoin C, Meunier PJ, Breart G. Separate and Combined Value of Bone Mass and Gait Speed Measurements in Screening for Hip Fracture Risk: Results from the EPIDOS Study. *Osteoporos Int* 9: 188-192; 1999
20. Davis JW, Ross PD, Wasnich RD, Maclean CJ, Vogel JM. Comparison of cross-sectional and longitudinal measurements of age-related changes in bone mineral content. *J Bone Miner Res* 4(3): 351-7; 1989
21. Denkschrift der Deutschen Forschungsgemeinschaft – Klinische Forschung. Schwartz FW, Badura B, Blanke B et al.(Hrsg.). VCH Verlagsgesellschaft Weinheim; 2000
22. Dretakis EK, Kontakis GM, Steriopoulos K, Dretakis K, Kouvidis G. Broadband ultrasound attenuation of the os calcis in female postmenopausal patients with cervical and trochanteric fractures. *Calcif Tissue Int* 57: 419-21; 1995
23. Dretakis KE, Dretakis EK, Papakitsou EF, Psarakis. Possible predisposing factors for the second hip fracture. *Calcif Tissue Int* 62: 366-369; 1998
24. Duppe H, Gardsell P, Nilsson B, Johnell O. A single density measurement can predict fractures over 25 years. *Calcif Tissue Int* 60: 171-4; 1997
25. Earnshaw SA, Hosking DJ. Clinical usefulness of risk factors for osteoporosis. *Ann Rheum Dis* 55(6): 338-9;1996
26. Einhorn TA. Bone strength, the bottom line (editorial) *Calcif Tissue Int* 51: 333-339; 1992
27. Ettinger B, Genant HK, Cann CE: Menopausal Bone Loss can be prevented by low dose estrogen with calcium supplements. *J Comput Assist Tomogr* 9: 633; 1985
28. Frost ML, Blake GM, Fogelman I. Contact quantitative ultrasound: an evaluation of precision, fracture discrimination, age-related bone loss and applicability of the WHO criteria. *Osteoporos Int* 10: 441-449;1999
29. Frost ML, Blake GM, Fogelman I. Does quantitative ultrasound imaging enhance precision and discrimination? *Osteoporos Int* 11: 425-433; 2000
30. Frost ML, Blake GM, Fogelman I. Can the WHO criteria for diagnosing osteoporosis be applied to calcaneal quantitative ultrasound? *Osteoporos Int* 11: 321-330; 2000
31. Garnerio P. Do markers of bone resorption add to bone mineral density and ultrasonographic heel measurement for the prediction of hip fracture in elderly women? The EPIDOS prospective study. *Osteoporos Int* 8: 563-569; 1998
32. Genant HK. Current state of bone densitometry for osteoporosis. *Radiographics*;18: 913-918; 1998

33. Glüer CC, Wu CY, Jergas M, Goldstein SA, Genant HK. Three quantitative ultrasound parameters reflect bone structure. *Calcif Tissue Int* 55: 46-52; 1994
34. Glüer CC, Cummings SR, Bauer DC. Osteoporosis: association of recent fractures with quantitative US findings. *Radiology* 199(3): 725-732; 1996
35. Glüer CC, for the International Quantitative Ultrasound Consensus Group. Quantitative ultrasound techniques for the assessment of osteoporosis: expert agreement on current status. *J Bone Miner Res* 12: 1280-8; 1997
36. Glüer CC, Barkmann R, Heller M. Ultraschall-schon praxisreif? *Therapeutische Umschau*, 55:667-75; 1998
37. Glüer CC, Barkmann R, Heller M. Quantitativer Ultraschall. Status 1999. *Radiologe* 39:213-21;1999
38. Glüer CC. Fortbildungshandbuch Osteoporose 2000, Berlin; 2000
39. Glüer CC, Barkmann R, Heller M. Osteoporosediagnostik mit Hilfe quantitativer Ultraschallverfahren. *Z.ärztl.Fortbild.Qual.sich.(ZaeFQ)* 94: 461-468; 2000
40. Glüer CC, Barkmann R. Aktuelle Kontroversen in der Osteoporosediagnostik. *Osteoporose*:6-10;2001
41. Gnudi S, Ripamonti C, Malavolta N. Quantitative Ultrasound and Bone Densitometry to Evaluate the Risk of Nonspine Fractures: A prospective Study. *Osteoporos Int* 11: 518-523; 2000
42. Gonnelli S, Cepollaro C, Agnusdei D. Diagnostic value of ultrasound analysis and bone densitometry as predictors of vertebral deformity in postmenopausal women. *Osteoporos Int* 5(6): 413-8;1995
43. Graafmans WC, Van Lingen A, Ooms ME. Ultrasound measurements in the calcaneus: precision and its relation with bone mineral density of the heel, hip and lumbar spine. *Bone* 19(2): 97-100; 1996
44. Grampp S, Henk CB, Imhof H. *Wiener Medizinische Wochenschrift Themenheft: "Osteoporose"* 194: 472-478;1999
45. Grampp S. Quantitative US of the calcaneus: cutoff levels for the distinction of healthy and osteoporotic individuals. *Radiology* 220: 400-405; 2001
46. Grampp S., Dobnig H, Willvonseder R, Leb G. Leitlinien zur Anwendung densitometrischer Verfahren. *J Miner Stoffwechs* 2: 50-51; 2001
47. Greenspan SL, Bouxsein ML, Melton ME. Precision and discriminatory ability of calcaneal bone assessment technologies. *J Bone Miner Res* 12 (8): 1303-13;1997
48. Gregg EW, Pereira MA, Caspersen CJ. Physical activity, falls and fractures among older adults: a review of the epidemiologic evidence. *J Am Geriatr Soc* 48(8): 883-93; 2000
49. Hans D, Dargent-Molina P, Schott AM. Ultrasonographic heel measurements to predict hip fracture in elderly women: the EPIDOS prospective study. *Lancet* 348(9026): 511-4;1996

50. Hans D, Wu C, Njeh CF. Ultrasound velocity of trabecular cubes reflects mainly bone density and elasticity. *Calcif Tissue Int* 64: 18-23; 1999
51. Heaney RP, Avioli LV, Chesnut III CH, Lappe J, Recker RR, Brandenburger GH. Osteoporotic bone fragility: detection by ultrasound transmission velocity. *JAMA* 261: 2986-90; 1989
52. Heaney RP. Pathophysiology of Osteoporosis. *Endocrinol Metab Clin North Am* 27(2): 255-265; 1998
53. He YQ. Assessment of a new quantitative ultrasound calcaneus measurement: Precision and discrimination of hip fractures in elderly women compared with dual X-ray absorptiometry. *Osteoporos Int* 11: 354-360; 2000
54. Heilbrun LK, Kagan A, Nomura A, Wasnich RD. The origins of epidemiologic studies of heart disease, cancer and osteoporosis among Hawaii Japanese. *Hawaii Med J* 44: 294-296; 1985
55. Harris S, Dawson Hughes B. Rates of changes in bone mineral density of the spine, heel, femoral neck and radius in healthy postmenopausal women. *Bone Miner* 17(1): 87-95; 1992
56. Homik J, Hailey D. Quantitative ultrasound for bone density measurement, Alberta Heritage Foundation for Medical Research, Canada, HTA 11; September 1998
57. Huang C, Ross PD, Yates AJ, Walker RE, Imose K, Emi K, Wasnich RD. Prediction of Fracture Risk by Radiographic Absorptiometry and Quantitative Ultrasound: A prospective Study. *Calcif Tissue Int* 63: 380-384; 1998
58. Iki M. Precision of quantitative ultrasound measurement of the heel bone and effects of ambient temperature on the parameters. *Osteoporos Int* 10:462-467;1999
59. Kang C. Comparison of ultrasound and dual energy X-ray absorptiometry measurements in the calcaneus. *Br J Radiol* 71: 861-867; 1998
60. Kanis J, Glüer CC, for the Committee of Scientific Advisors. IOF: an update on the diagnosis and assessment of osteoporosis with densitometry. *Osteoporos Int* 11: 192-202; 2000
61. Kann PH. Osteodensitometrie und Ultraschalluntersuchungen des Knochens. Methoden, Indikationen und Aussagekraft. *Orthopäde* 30: 437-443; 2001
62. Katz J, Meunier A. The elastic anisotropy of bone. *J Biomechanics* 20: 1063-70; 1987
63. Köbberling J, Richter K, Trampisch H-J, Windeler J: Methodologie der medizinischen Diagnostik. Entwicklung, Beurteilung und Anwendung von Diagnoseverfahren in der Medizin. Springer-Verlag; 1991
64. Kolta S, Fechtenbaum J, Roux C. Bone assessment using a contact ultrasound system. *J Bone Miner Res* 12(1)S386; 1997
65. Lange S, Richter K, Köbberling J. Die Osteodensitometrie – Metaanalyse über den diagnostischen Wert bei der Osteoporose. Lit-Verlag Münster – Hamburg; 1994
66. Langton CM, Palmer SB, Porter RW. The measurement of broadband ultrasound attenuation in cancellous bone. *Engl Med* 13:89-91; 1984

67. Laugier P, Giat P, Berger G. Broadband ultrasonic attenuation imaging: a new imaging technique of the os calcis. *Calcif Tissue Int* 54: 83-86; 1994
68. Laugier P, Droin P, Laval-Jeantet AM. In vitro assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and quantitative computed tomography. *Bone* 20: 157-165; 1997
69. Lühmann D, Kohlmann T, Lange S, Raspe H. Die Rolle der Osteodensitometrie im Rahmen der Primär-, Sekundär- und Tertiärprävention/Therapie der Osteoporose. HTA 13, Nomos Verlagsgesellschaft, Baden-Baden; 2000
70. Machado A, Hannon R, Henry Y. Standardised coefficient of variation for dual energy x-ray absorptiometry (DXA), quantitative ultrasound (QUS) and markers of bone turnover. *J Bone Miner Res* 12(1)S 258; 1997
71. Marshall D, Johnell O, Wedel H. Meta analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. *BMJ* 312(7041): 1254-9; 1996
72. Matthis C, Raspe H und die EVOS-Gruppe Deutschland. Krankheitslast bei vertebralem Deformitäten. *Med Klin* 93 (II): 41-46; 1998
73. Mele R, Masci G, Ventura V, De Aloysio D, Bicocchi M, Cadossi R. Three-year longitudinal study with quantitative ultrasound at the hand phalanx in a female population. *Osteoporos Int* 7: 550-7; 1997
74. Melton III LJ, O'Fallon WM, Riggs BL: Secular Trends in the incidence of hip fractures. *Calcified Tissue International* 41: 57-64; 1987
75. Melton LJ III. Epidemiology of fractures. In: Riggs BL, Melton LJ III, eds.: *Osteoporosis: etiology, diagnosis, and management*. Philadelphia: Lippincott-Raven Publishers; 1995
76. Miller CG, Herd RJM, Ramalingam T, Fogelman I, Blake GM. Ultrasonic velocity measurements through the calcaneus: Which velocity should be measured? *Osteoporos Int* 3: 31-35, 1993
77. Morris R, Mazess R, Trempe J. Stiffness index compensates for temperature variation in ultrasound densitometry. *J Bone Miner Res* 12: (1) S 388; 1997
78. National Institutes of Health: *Osteoporosis prevention, diagnosis and therapy*. *JAMA* 285; 785-795; 2001
79. National Osteoporosis Foundation: *Guidelines Osteoporosis*. *Osteoporos Int* 7: 80; 1998
80. Nevitt MC, Johnell O, Black DM, Ensrud K, Genant HK, Cummings SR. Bone mineral density predicts non-spine fractures in very elderly women. Study of Osteoporotic Fracture Research Group. *Osteoporos Int* 4: 325-31; 1994
81. Nguyen T, Sambrook P, Eisman JA: Sources of variability in bone mineral density measurements: implications for study design and analysis of bone loss. *J Bone Miner Res* 12: 124-135; 1993
82. Njeh CF, Boivin CM, Langton CM. The role of ultrasound in the assessment of osteoporosis: a review. *Osteoporos Int* 7: 7-22; 1997

83. Njeh CF, Hans D, Fuerst T, Glüer CC, Genant HK. Quantitative Ultrasound; assessment of osteoporosis and bone status. Martin Dunitz, London; 1999
84. Njeh CF. Association between QUS and structural parameters? *Calcif Tissue Int* 67: 273-274; 2000
85. Palmer SB, Langton CM. Ultrasonic studies of the os calcis. In: *In vivo Body Composition Studies*, (eds Ellis KJ, Yasumura S, Morgan WD): 452-46; 1997
86. Pfeifer M, Minne HW. Schenkelhalsfrakturen in Deutschland. Prävention, Therapie, Inzidenz und sozioökonomische Bedeutung. *Dt.Ärzteblatt* 26 (98A): 1751-1757; 2001
87. Pluijm SMF, Graafmans WC, Bouter LM. Ultrasound measurements for the prediction of osteoporotic fractures in elderly people. *J Bone Miner Res* 12 (1): S 362; 1997
88. Pluijm SMF, Graafmans WC, Bouter LM, Lips P. Ultrasound Measurements for the Prediction of Osteoporotic Fractures in Elderly People. *Osteoporos Int* 9: 550-556; 1999
89. Porter RW, Miller CG, Grainger D. Prediction of hip fracture in elderly women: a prospective study. *BMJ* 301: 638-41; 1990
90. Prins SH, Jorgensen HL, Jorgensen LV, Hassager C.. The role of quantitative ultrasound in the assessment of bone: a review. *Clin Physiol* 18: 3-17; 1997
91. Riggs BL, Melton LJ. The worldwide problem of osteoporosis: Insights afforded by epidemiology. *Bone* 17: 505s-511s; 1995
92. Ross P, Huang C, Davis J. Predicting vertebral deformity using bone densitometry at various skeletal sites and calcaneus ultrasound. *Bone* 16 (3): 325-32; 1995
93. Roux C, Fournier B, Laugier P, Chappard C, Kolta S, Dougados M, Berger G. Broadband ultrasound attenuation imaging: a new imaging method in osteoporosis. *J Bone Miner Res* 11: 1112-1118; 1996
94. Sahota O. A comparison of the longitudinal changes in quantitative ultrasound with dual-energy X-ray absorptiometry: the four-year effects of hormone replacement therapy. *Osteoporos Int* 11: 52-58; 2000
95. Schott AM, Weill-Engerer S, Hans D, Dubouef F, Delmas PD, Meunier PJ. Ultrasound discriminates patients with hip fracture equally well as dual energy X-ray absorptiometry and independently of bone mineral density. *J Bone Miner Res* 10: 243-9; 1995
96. Schürch MA, Rizzoli R, Mermillod B, Vasey H, Michel JP, Bonjour JP. A prospective study on socioeconomic aspects of fracture of the proximal femur. *J Bone Miner Res* 12: 1935-1942; 1996
97. Shutilov V: Physik des Ultraschalls. Springer Verlag, Wien, New York; 1984
98. Steckelberg A, Balgenorth A, Mühlhauser I. Analyse von deutschsprachigen Verbraucher-Informationsbroschüren zum Screening auf kolorektales Karzinom. *Z.ärztl.Fortbild.Qual.sich.(ZaeFQ)* 95: 535-538; 2001
99. Stewart A, Reid DM, Porter RW. Broadband ultrasound attenuation and dual energy X-ray absorptiometry in patients with hip fractures: Which technique discriminates fracture risk. *Calcif Tissue Int* 54(6): 466-9; 1994

100. Stewart A, Torgerson DJ, Reid DM. Prediction of fractures in perimenopausal women: A comparison of dual energy x-ray absorptiometry and broadband ultrasound attenuation. *Ann Rheum Dis* 55(2): 140-2; 1996
101. Stewart A, Walker LG, Porter RW, Reid DM, Primrose WR. Predicting a second hip fracture. *J Clin Densitom* 2: 363-370; 1999
102. Strelitzki R, Clarke AJ, Truscott JG. Ultrasonic measurement: an evaluation of three heel bone scanners compared with a bench-top system. *Osteoporos Int* 6(6): 471-9; 1996
103. Thompson P, Taylor J, Fisher A. A prospective study of fracture prediction using heel ultrasound in postmenopausal women. *J Bone Miner Res* 11: 1829; 1996
104. Turner CH, Peacock M, Timmermann L. Calcaneal ultrasonic measurements discriminate hip fracture independently of bone mass. *Osteoporos Int* 5(2): 130-5; 1995
105. van Daele PL, Burger H, Hofman A. Ultrasound assessment of the calcaneus. For the time being a conservative approach in clinical use. The ERGO Study. *Ned Ned Tijdschr Geneesk* 140(27): 1398-1402; 1996
106. Wasnich RD, Ross PD, Davis JW, Vogel JM. A comparison of single and multi-site BMC measurements for assessment of spine fracture probability. *J Nucl Med* 30: 1166-71; 1989
107. WHO. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: a report of a WHO study group, in World Health Organization, Geneva, Switzerland. WHO Technical Report Series, Geneva: 843; 1994
108. Wildner M, Caspar W, Bergmann KE. Ein probabilistisches Modell zur Abschätzung des Lebenszeitrisikos osteoporotischer Frakturen. Poster, 41. Jahrestagung der GMDS, Bonn 15-19.9.1996
109. Wüster C. Osteosonographie – praktische Anwendung und zukünftige Entwicklungen. *Med-Report* 2, Osteologie-Kongreß 2/1999
110. Wüster C, Albanese C, de Aloysio D. Phalangeal osteosonogrammetry study: age-related changes, diagnostic sensitivity, and discrimination power. *J Bone Miner Res* 15: 1603-1614; 2000

ANHANG

zu SOP 2: Datenbankbeschreibungen

Name	MEDLINE
Hersteller	National Library of Medicine, USA
Anschrift	Building 38, Mail Stop 20, 8600 Rockville Pike, Bethesda, MD 20894
Internetadresse	http://gateway.nlm.nih.gov/ oder www.dimdi.de
Inhaltsbeschreibung	Entspricht inhaltlich dem Index Medicus und einigen anderen gedruckten Bibliographien. Fachgebiete: Gesamte Medizin, einschließlich Zahn- und Veterinärmedizin. Außerdem Randgebiete wie z.B. Biologie, Biochemie, Biophysik und Psychologie.
Suchmöglichkeiten	Thesaurus: Ja (MeSH) Index: Ja Feldsuchen: Ja
Name	HEALTHSTAR
Hersteller	National Library of Medicine, USA
Anschrift	Building 38, Mail Stop 20, 8600 Rockville Pike, Bethesda, MD 20894
Internetadresse	http://gateway.nlm.nih.gov/ oder http://www.dimdi.de
Inhaltsbeschreibung	Fachgebiete: Nicht-klinische Aspekte des Gesundheitswesens, insbesondere Planung, Organisation, Verwaltung, Finanzierung, Ausbildung, Personal- und Versicherungsfragen sowie Qualitätssicherung und Bewertung. Aus MEDLINE werden alle Zitate zu diesem Thema in HEALTHSTAR übernommen. Darüberhinaus werden weitere Zeitschriften, Bücher, Konferenz- und Regierungsberichte berücksichtigt.
Suchmöglichkeiten	wie Medline
Name	The COCHRANE Library
Hersteller	Update Software Ltd, CD-ROM oder Online Version; ca. 130 britische Pfund / Anno / Einzelplatzlizenz
Anschrift	Summertown Pavilion Middle Way, Oxford OX2 7LG, UK; Tel:+44 1865 513902, Fax:+44 1865 516918 oder Update Software Inc,936 La Rueda Road, Vista, California 92084 USA Tel:+1 760 727-6792, Fax:+1 760 734-4351
Internetadresse	http://www.update-software.com ; http://www.cochrane.co.uk
Inhaltsbeschreibung	Die Cochrane Library enthält die Reviews der Cochrane Collaboration (CDSR), die Datenbank DARE, das Cochrane Register für kontrollierte Studien (CCTR), eine HTA-Datenbank und eine gesundheitsökonomische Reviewdatenbank (EED)
Suchmöglichkeiten	Thesaurus: Ja (MeSH) Index: Ja Feldsuchen: Ja
Name	DARE
Hersteller	University of York (NHS Centre for Reviews and Dissemination (CRD))
Anschrift	NHS Centre for Reviews and Dissemination, University of York, YO1 5DD, United Kingdom
Internetadresse	http://nhscrd.york.ac.uk

Inhaltsbeschreibung	Die Datenbank enthält strukturierte Abstracts von Reviews, die nach einem bestimmten Kriterienkatalog für qualitativ hochwertig erachtet wurden
Suchmöglichkeiten	Thesaurus: Ja (enthält MeSH und NON-MeSH Begriffe (letzte sind gekennzeichnet)) Schlagwortsuche: Ja
Name	HSTAT (Health Services Technology Assessment Texts)
Hersteller	National Information Center on Health Services Research and Health Care Technology (NICHSR), National Library of Medicine (NLM),
Anschrift	Building 38, Mail Stop 20, 8600 Rockville Pike, Bethesda, MD 20894
Internetadresse	http://text.nlm.nih.gov
Inhaltsbeschreibung	Bei dieser Datenbank handelt es sich um ein Archiv elektronischer Texte zu dem Bereich Health Services Technology Assessment HSTAT enthält Guidelines, Quick Reference Guides, Verbraucher Broschüren, Evidence Reports, Technology-Assessment-Reports, Konsensusberichte u.v.a.
Suchmöglichkeiten	nur Freitext
Name	HTA Database
Hersteller	University of York (NHS Centre for Reviews and Dissemination/CRD)
Anschrift	NHS Centre for Reviews and Dissemination, University of York, York, YO1 5DD, United Kingdom
Internetadresse	http://nhscrd.york.ac.uk
Inhaltsbeschreibung	Datenbank, die strukturierte Abstracts von HTA-Reports verschiedener internationaler HTA-Institutionen enthält
Suchmöglichkeiten	Thesaurus: Ja (MeSH und NON-MeSH-Begriffe, die als solche gekennzeichnet sind)
Name	NEED (NHE Economic Evaluation database)
Hersteller	University of York (NHS Centre for Reviews and Dissemination/CRD)
Anschrift	NHS Centre for Reviews and Dissemination, University of York, York, YO1 5DD, United Kingdom
Internetadresse	http://nhscrd.york.ac.uk
Inhaltsbeschreibung	Datenbank mit strukturierten Abstracts im Bereich Gesundheitsökonomie sowie Bibliographien zu ausgewählten Themen -
Suchmöglichkeiten	Thesaurus: Ja (MeSH und NON-MeSH-Begriffe; letztere sind gekennzeichnet)
Name	ISTAHC Database
Hersteller	ISTAHC (International Society of Technology Assessment in Health Care)
Anschrift	380 St. Antoine West, Suite 3200, Montreal (Quebec) H2Y 3X7, CANADA / Tel:(514)844-3033, Fax:(514)987-1567
Internetadresse	http://www.istahc.org
Inhaltsbeschreibung	Die Datenbank enthält Literatur zum Thema HTA in Form von Reviews, Konferenzberichten, Konferenzprogrammen und Artikeln. Als Quelle dienen dabei die Publikationen der ISTAHC, beispielsweise das International Journal of Technology Assessment in Health Care
Suchmöglichkeiten	nur Freitext

zu SOP 3: Formulare der German Scientific Working Group

Checkliste 1a: Kontextdokumente				
Bericht Nr.:				
Titel:				
Autoren:				
Quelle:				
Dokumenttyp HTA-Bericht <input type="checkbox"/> Praxisrichtlinie <input type="checkbox"/> Anderes Dokument <input type="checkbox"/>				
Adressaten: Entscheidungsträger <input type="checkbox"/> Kliniker <input type="checkbox"/> Patienten <input type="checkbox"/> Andere <input type="checkbox"/>				
Klas	A Fragestellung und Kontext	Ja	Nein	?
I	1. Werden Anlass und Ziel der Publikation im Sinne einer "Policy Question" dargestellt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	2. Gibt es im Rahmen des breiteren Kontext eine präzise formulierte Forschungsfrage?*	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	3. Sind in der Publikation Angaben zu folgenden Aspekten enthalten:	<input type="checkbox"/>	<input type="checkbox"/>	
I	a) Epidemiologie der Zielerkrankung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	b) (Entwicklungs-)stand der Technologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	c) Efficacy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	d) Effectiveness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	e) Nebenwirkungen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	f) Indikationen**	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	g) Kontraindikationen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	h) Praxisvariation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	i) Versorgungsstrukturen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	j) Kostengesichtspunkten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	k) sozioökonomischem, ethischem und juristischem Impact	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Klas	B Methodik der Informationsgewinnung			
QA	1. Wurden die genutzten Quellen dokumentiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	2. Wurden die Suchstrategien dokumentiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	3. Wurden Einschlusskriterien definiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	4. Wurden Ausschlusskriterien definiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Klas	C Methodik der Bewertung und Dokumentation:			
QA	1. Wurden Validitätskriterien berücksichtigt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QC	2. Wurde die Bewertung unabhängig von mehreren Personen durchgeführt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QC	3. Sind ausgeschlossene Studien mit ihren Ausschlussgründen dokumentiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QC	4. Ist die Datenextraktion nachvollziehbar dokumentiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QC	5. Erfolgte die Datenextraktion von mehreren Personen unabhängig?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Klas	D Methodik der Informationssynthese			
I	1. Es wurden quantitative Informationssynthesen durchgeführt:(bitte für die enthaltene Meta-Analyse Bogen 1b ausfüllen)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	2. Es wurden qualitative Informationssynthesen durchgeführt: (dann bitte auch Bogen 1b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	3. Es wurden zur Ergänzung der Datenlage eigene Erhebungen durchgeführt:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E Ergebnisse / Schlussfolgerungen				
QB	1. Wird die bestehende Evidenz in den Schlussfolgerungen konsequent umgesetzt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	2. Werden methodisch bedingte Limitationen der Aussagekraft kritisch diskutiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	3. Werden Handlungsempfehlungen ausgesprochen? ***	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	4. Gibt es ein Grading der Empfehlungen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QC	5. Wurde die Publikation vor der Veröffentlichung einem externen Reviewverfahren unterzogen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	6. Ist ein "Update der Publikation eingeplant?"	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	3. Qualitative Informationssynthesen:			
QA	a) Ist die Informationssynthese nachvollziehbar dokumentiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	b) Gibt es eine Bewertung der bestehenden Evidenz?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E Schlussfolgerungen				
QB	1. Wird die Forschungsfrage beantwortet?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	2. Wird die bestehende Evidenz in den Schlussfolgerungen konsequent umgesetzt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	3. Werden methodisch bedingte Limitationen der Aussagekraft kritisch diskutiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	4. Werden Handlungsempfehlungen ausgesprochen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	5. Gibt es ein Grading der Empfehlungen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	5. Wird weiterer Forschungsbedarf identifiziert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	6. Ist ein "Update" des Review eingeplant?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Abschließende Beurteilung: Die vorliegende Publikation wird berücksichtigt <input type="checkbox"/> ausgeschlossen <input type="checkbox"/>				

Checkliste 2a: Primärstudien (RCTs / Fall-Kontrollstudien / Kohortenstudien / Längsschnittstudien / Fallserien)				
Bericht Nr.:				
Titel:				
Autoren:				
Quelle:				
Dokumenttyp	RCT: <input type="checkbox"/>	Kohortenstudie: <input type="checkbox"/>	Fall-Kontrollstudie: <input type="checkbox"/>	Längsschnittstudie: <input type="checkbox"/>
	Fallserie: <input type="checkbox"/>	Andere: <input type="checkbox"/>		
Klas	A Auswahl der Studienteilnehmer	Ja	Nein	?
QA	1. Sind die Ein- und Ausschlusskriterien für Studienteilnehmer ausreichend / eindeutig definiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	2. Wurden die Ein-/ Ausschlusskriterien vor Beginn der Intervention festgelegt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	3. Wurde der Erkrankungsstatus valide und reliabel erfaßt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QBI	4. Sind die diagnostischen Kriterien der Erkrankung beschrieben?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	5. Ist die Studienpopulation / exponierte Population repräsentativ für die Mehrheit der exponierten Population bzw. die "Standardnutzer" der Intervention?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	6. Bei Kohortenstudien: Wurden die Studiengruppen gleichzeitig betrachtet?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	B Zuordnung und Studienteilnahme			
QA	1. Entstammen die Exponierten / Fälle und Nicht-Exponierten / Kontrollen einer ähnlichen Grundgesamtheit?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	2. Sind Interventions-/Exponierten- und Kontroll-/ Nicht-Exponiertengruppen zu Studienbeginn vergleichbar?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	3. Erfolgte die Auswahl randomisiert mit einem standardisierten Verfahren?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QC	4. Erfolgte die Randomisierung blind?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	5. Sind bekannte / mögliche Confounder zu Studienbeginn berücksichtigt worden?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	C Intervention / Exposition			
QA	1. Wurden Intervention bzw. Exposition valide, reliabel und gleichartig erfaßt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	2. Wurden Interventions- / Kontrollgruppen mit Ausnahme der Intervention geichartig therapiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	3. Falls abweichende Therapien vorlagen, wurden diese valide und reliabel erfaßt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	4. Bei RCTs: Wurden für die Kontrollgruppen Placebos verwendet?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	5. Bei RCTs: Wurde dokumentiert wie die Plazebos verabreicht wurden?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	D Studienadministration			
QB	1. Gibt es Anhaltspunkte für ein "Overmatching"?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	2. Waren bei Multicenterstudien die diagnostischen und therapeutischen Methoden sowie die Outcome-Messung in den beteiligten Zentren identisch?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

E Outcome Messung				
I	1. Wurden patientennahe Outcome-Parameter verwendet?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	2. Wurden die Outcomes valide und reliabel erfaßt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	3. Erfogte die Outcome Messung verblindet?			
QC	4. Bei Fallserien: Wurde die Verteilung prognostischer Faktoren ausreichend erfaßt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F Drop Outs				
QA	4. War die Response-Rate bei Interventions-/ kontrollgruppen ausreichend hoch bzw. bei Kohortenstudien: konnte ein ausreichend großer Teil der Kohorte über die gesamte Studiendauer verfolgt werden?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	5. Wurden die Gründe für Ausscheiden von Studienteilnehmern aufgelistet?			
QB	6. Wurden die Outcomes der Drop-Outs beschrieben und in der Auswertung berücksichtigt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	7. Falls Differenzen gefunden wurden - sind diese signifikant?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	8. Falls Differenzen gefunden wurden - sind diese relevant?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
G Statistische Analyse				
QA	1. Sind die beschriebenen analytischen Verfahren korrekt und die Informationen für eine einwandfreie Analyse ausreichend?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	2. Wurden für Mittelwerte und Signifikanztests Konfidenzintervalle angegeben?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	3. Sind die Ergebnisse in graphischer Form präsentiert und wurden die den Graphiken zugrundeliegenden Werte angegeben?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Beurteilung:				
Die vorliegende Publikation wird:				
berücksichtigt <input type="checkbox"/>				
ausgeschlossen <input type="checkbox"/>				
Legende:				
Klass. Klassifikation der Frage				
Q Frage, die Aspekte der methodischen Qualität erfasst; in absteigender Relevanz mit A, B oder C bewertet				
I Frage, mit reinem Informationsgehalt, irrelevant für die Qualitätsbeurteilung				

Checkliste 2b: Diagnosestudie				
Bericht Nr.:				
Titel:				
Autoren:				
Quelle:				
Klas	A Beschreibung der Ausgangssituation	Ja	Nein	?
QA	1. Gibt es eine klar formulierte Fragestellung vor Beginn der Studie?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	2. Wurde die Zielkrankheit eindeutig definiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	3. Erfolgte die Festlegung der Trenngröße vor Beginn?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	4. Wurde ein "Goldstandard" festgelegt Angaben über seine Zuverlässigkeit gemacht?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B Durchführung der Prüfung				
QB	1. Ausreichende Beschreibung für Nachprüfungen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	2. Erfolgte die Auswertung des Testergebnisses ohne Kenntnis der Diagnose und umgekehrt die Diagnose ohne Kenntnis des Testergebnisses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	3. Wurde die Zusammensetzung der Versuchskollektive in bezug auf die Übertragbarkeit der Ergebnisse berücksichtigt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QA	Wurde die zu untersuchende Technik und der "Goldstandard" bei allen Patienten angewendet?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C Ergebnisinterpretation				
QA	1. Ist eine Vierfeldertafel vorhanden bzw. ist eine Erstellung aus den gegebenen Daten möglich?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	2. Erfogte eine Verwendung von eindeutig definierten Parametern zur Beschreibung der Ergebnisse?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	D Diskussion			
QB	1. Wurde die Abhängigkeit des prädiktiven Wertes von der a priori Wahrscheinlichkeit ausreichend diskutiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	2. Wurde die Definition des pathologischen Testergebnisses als Erkrankung selbst vermieden?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	3. Wurde eine Nutzen-Schaden-Abwägung für die vier Ergebnisgruppen durchgeführt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
QB	4. Wurden die Folgen von Fehlklassifikationen angemessen diskutiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Beurteilung:				
Die vorliegende Publikation wird: berücksichtigt <input type="checkbox"/> ausgeschlossen <input type="checkbox"/>				
Legende:				
Klass. Klassifikation der Frage				
Q	Frage, die Aspekte der methodischen Qualität erfasst; in absteigender Relevanz mit A, B oder C bewertet			
I	Frage, mit reinem Informationsgehalt, irrelevant für Qualitätsbeurteilung			

SOP 6: Levels of Evidence für Prognosestudien nach CEBM (2001)

Tabelle 36: Evidenzlevel nach CEBM (2001) für Prognosestudien

Level	Prognosis
1a	SR (with <u>homogeneity*</u>) of inception cohort studies; <u>CDR†</u> validated in different populations
1b	Individual inception cohort study with ≥ 80 % follow-up; <u>CDR†</u> validated in a single population
1c	All or none case-series
2a	SR (with <u>homogeneity*</u>) of either retrospective cohort studies or untreated control groups in RCTs
2b	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of <u>CDR†</u> or validated on split-sample ^{§§§} only
2c	"Outcomes" Research
3a	-
3b	-
4	Case-series (and <u>poor quality prognostic cohort studies***</u>)
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"
Level	Prognosis
*	By homogeneity we mean a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant. As noted above, studies displaying worrisome heterogeneity should be tagged with a "-" at the end of their designated level.
†	Clinical Decision Rule. (These are algorithms or scoring systems which lead to a prognostic estimation or a diagnostic category).
§§§	Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into "derivation" and "validation" samples.
***	By poor quality prognostic cohort study we mean one in which sampling was biased in favour of patients who already had the target outcome, or the measurement of outcomes was accomplished in < 80 % of study patients, or outcomes were determined in an unblinded, non-objective way, or there was no correction for confounding factors.

Kooperationspartner

Personen aus folgenden Institutionen und Fachabteilungen wurden um Kommentare und Rückmeldung im Rahmen der Erstellung von SOP 1 gebeten:

Potentielle Nutzer:

Kassenärztliche Bundesvereinigung (KBV), Medizinischer Dienst der Krankenkassen (MDK), Medizinischer Dienst der Spitzenverbände (MDS)

Methodische Experten:

Universitätsabteilung für Epidemiologie, Sozialmedizin und Gesundheitssystemforschung; Deutsches Institut für Medizinische Dokumentation und Information (DIMDI)

Experten für Krankheitsbild und Technologie:

Universitätsabteilung für Diagnostische Radiologie, Osteologische Arbeitsgruppe

Prä- und Post-Delphi-Verfahren

Der nachstehende Fragenkatalog wurde im Rahmen eines modifizierten Delphi-Verfahrens an den vorher beschriebenen Personenkreis geschickt. Es wurde eine Befragungsrunde durchgeführt, sowie nach Zusammenfassung der unterschiedlichen Meinungen und Kommentare eine Abschlussrunde:

1. Wie korrekt lässt sich das Frakturrisiko mithilfe der quantitativen Ultraschallverfahren abschätzen (relatives Frakturrisiko im Vergleich zu Personen mit Normalwerten)?
2. Welches der unterschiedlichen QUS-Verfahren ist hierzu am besten geeignet?
3. Welche qualitätssichernden Maßnahmen sind vor einem Einsatz in der ärztlichen Praxis zu realisieren, um zu validen Ergebnissen zu kommen?
4. Gibt es weiteren Forschungsbedarf zur Klärung des Stellenwertes der Ultraschalldiagnostik im Rahmen der Osteoporoseversorgung?
5. Wie wären die ermittelten Ergebnisse zur Patienteninformation und Entscheidungshilfe nutzbar?

Im Folgenden findet sich die überarbeitete Fassung des Fragenkataloges, der schlussendlich im HTA-Schnellverfahren bearbeitet wurde:

1. Wie valide lässt sich das Frakturrisiko mithilfe der quantitativen Ultraschallverfahren im Vergleich zum DXA-Verfahren abschätzen (prädiktive Validität: relatives Frakturrisiko im Vergleich zu Personen mit Normalwerten)?
2. Welches der unterschiedlichen QUS-Verfahren ist hierzu am besten geeignet?
3. Sollte die Analyse ergeben, dass die quantitative Ultraschalldiagnostik eine valide und zuverlässige Abschätzung des Frakturrisikos erlaubt: Welche qualitätssichernden Maßnahmen sind dann vor einem Einsatz in der ärztlichen Praxis zu realisieren, um zu reliablen und unverzerrten Ergebnissen zu kommen?
4. Welchen weiteren Forschungsbedarf zur Klärung des Stellenwertes der Ultraschalldiagnostik im Rahmen der Frakturprävention gibt es?
5. Wie wären die unter 1 – 3 ermittelten Ergebnisse zur Patienteninformation und zum Verbraucherschutz zu nutzen?

Datenbankrecherchen (Primäre Recherche)

Ein- und Ausschlusskriterien:

Die Titel der Zitate sowie die Abstracts wurden von 2 unabhängigen Untersuchern manuell nach den folgenden Ein- und Ausschlusskriterien auf potentielle Relevanz für das HTA-Vorhaben selektiert:

Tabelle 37: Ein- und Ausschlusskriterien: QUS

Relevanz	Kriterien
zentral	Ultraschallmessung, Frakturrisiko, prospektive Studie
peripher	Ultraschall und DXA, Querschnittsstudie, Ultraschall, populationsbezogene Stichprobe: Normwerte, Frakturpatienten im Vergleich zu Personen ohne Frakturen
Ausschluss	Sekundäre Osteoporosen; in vitro Studien, Tierstudien, thematisch abweichend, Briefe, Editorials, Fallberichte; Sprache nicht Englisch, Deutsch, Französisch, Spanisch; eindeutig unsystematische Übersichtsartikel
unklar	anhand der Angaben im Abstract nicht eindeutig zuzuordnen

Recherchierte Datenbanken:

Tabelle 38: Dokumentationsformular: Datenbank, QUS

Datenbankname	Zugriff	Recherchierte Jahrgänge
Cochrane Library	CD Version, Issue 4, 2000	bis 12/2000
ISTAHC-Datenbank	ISTAHC Website (www.isthac.org)	letzter Zugriff 2/2001
Medline	DIMDI	1966-2/2001
Healthstar	DIMDI	1975-2/2001

Suchstrategien

a) Medline und Healthstar

Tabelle 39: Dokumentationsformular: Recherche, Medline + Healthstar

No.	Records	Request
# 1	14287533	HE 75; ME 66
# 2	24737	CT D osteoporosis
# 3	13450	CT D hip fractures
# 4	4706	CT D spinal fractures
# 5	5290	CT D radius fractures

# 6	19036	CT D bone density
# 7	3788	Osteopen?
# 8	31060	Osteopor?
# 9	1444	Fractur?risk
# 10	5250	Spina?fractur?
# 11	2176	Vertebr?fractur?
# 12	4987	Radius fractur?
# 13	9105	Hip fractur?
# 14	1045	Colles fractur?
# 15	21102	Bone density
# 16	64966	S=15 OR S=14 OR S=13 OR S=12 OR S=11 OR S=10 OR S=9 OR S=8 OR S=7 OR S=6 OR S=5 OR S=4 OR S=3 OR S=2
# 17	3242	SOS
# 18	534	BUA
# 19	238	QUS
# 20	101637	Sonograph?
# 21	88896	Ultrasound
# 22	11	Osteosonograph?
# 23	192239	Ultrasonograph?
No.	Records	Request
# 24	239312	S=24 OR S=23 OR S=22 OR S=21 OR S=20 OR S=19 OR S=18
# 25	1607	S=24 AND S=16
# 26	915	Unique in 25

b) Cochrane Library (Issue 4, 2000)

In der Cochrane Library wurde mit der folgenden Strategie nach systematischen Übersichten und HTA-Berichten zur Thematik recherchiert.

Tabelle 40: Dokumentationsformular: Recherche, Cochrane Library

No.	Records	Request
# 1	1033	BONE-DENSITY*:ME
# 2	1324	(BONE next DENSITY)
# 3	347	HIP-FRACTURES*:ME
# 4	89	RADIUS-FRACTURES*:ME
# 5	100	SPINAL-FRACTURES*:ME
# 6	486	(HIP next FRACTUR*)
# 7	115	(COLLES next FRACTUR*)
# 8	133	(RADI* next FRACTUR*)

# 9	189	(VERTEBR* next FRACTUR*)
# 10	141	(SPIN* next FRACTUR*)
# 11	61	(FRACTUR* next RISK)
# 12	857	OSTEOPOROSIS*.ME
# 13	1656	OSTEOPOR*
# 14	150	OSTEOPEN*
# 15		#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14
# 16	2415	ULTRASOUND
# 17	3149	ULTRASONOGRAPH*
# 18	8	QUS
# 19	15	BUA
# 20	39	SOS
# 21	583	SONOGRAPH*
# 22	0	OSTEOSONOGRAPH*
# 23	4783	#16 or #17# or #18 or #19 or #20 or #21 Or #22
# 24	112	#15 and #23

c) ISTAHC

In der Datenbank der International Society for Technology Assessment wurden unter den Stichworten „Osteoporosis“ und „Ultrasound“ insgesamt 30 Treffer erzielt.

Eine Publikation (HTA-Bericht von Homik J & Hailey D, September 1998) wurde in unsere Auswertung einbezogen.

Selektion

Der HTA-Bericht aus dem Jahr 1998 wird einer vorläufigen Qualitätsbewertung unterzogen: Er enthält eine systematische Zusammenfassung des adäquaten Studientyps; somit kann die Literatur-Recherche auf die Jahrgänge 1998 – 2001 begrenzt werden. Alle Literaturstellen ≤ **1997** werden zurückgestellt.

Nach einer 1. Selektion anhand von Abstracts verbleiben 126 Publikationen, nach einer 2. Selektion anhand der vollständigen Artikel verbleiben noch 39 Literaturstellen (von ursprünglich 916), d. h. aufgeschlüsselt:

- 7 Prognosestudien
- 30 Phase-II-Studien (diagnostisch)
- 1 systematischer Review
- 1 HTA-Bericht (Homik/Hailey 1998).

Liste der eingeschlossenen Publikationen (7 + 1 + 1), die anhand der neu entwickelten Stufenplänen/Checklisten für systematische Übersichtsarbeiten und Kohortenstudien bewertet werden:

1. HTA-Berichte:

1. Homik J, Hailey D. Quantitative ultrasound for bone density measurement. Alberta Heritage Foundation for Medical Research, Canada, HTA 11; September 1998

2. Systematische Reviews:

1. Prins SH, Jorgensen HL, Jorgensen LV, Hassager C. The role of quantitative ultrasound in the assessment of bone: a review. *Clin Physiol* 18: 3-17; 1997

3. Prognosestudien:

1. Dretakis KE, Dretakis EK, Papakitsou EF, Psarakis. Possible predisposing factors for the second hip fracture. *Calcif Tissue Int* 62: 366-369; 1998
2. Garnero P. Do markers of bone resorption add to bone mineral density and ultrasonographic heel measurement for the prediction of hip fracture in elderly women? The EPIDOS prospective study. *Osteoporos Int* 8: 563-569; 1998
3. Dargent-Molina P, Schott AM, Hand D, Favier F, Grandjean H, Baudoin C, Meunier PJ, Breart G. Separate and Combined Value of Bone Mass and Gait Speed Measurements in Screening for Hip Fracture Risk: Results from the EPIDOS Study. *Osteoporos Int* 9: 188-192; 1999
4. Gnudi S, Ripamonti C, Malavolta N. Quantitative Ultrasound and Bone Densitometry to Evaluate the Risk of Nonspine Fractures: A prospective Study. *Osteoporos Int* 11: 518-523; 2000
5. Huang C, Ross PD, Yates AJ, Walker RE, Imose K, Emi K, Wasnich RD. Prediction of Fracture Risk by Radiographic Absorptiometry and Quantitative Ultrasound: A prospective Study. *Calcif Tissue Int* 63: 380-384; 1998
6. Pluijm SMF, Graafmans WC, Bouter LM, Lips P. Ultrasound Measurements for the Prediction of Osteoporotic Fractures in Elderly People. *Osteoporos Int* 9: 550-556; 1999
7. Stewart A, Walker LG, Porter RW, Reid DM, Primrose WR. Predicting a second hip fracture. *J Clin Densitom* 2: 363-370; 1999

Von den Prognosestudien erfüllten fünf Publikationen (Nr. 3 - 7) die Einschlusskriterien.

Bewertung:

Bewertungsinstrumente für eingeschlossene Publikationen

Die Bewertungsinstrumente, mit denen die Qualitätsbeurteilung vorgenommen wurde, sind in Teil I dieser Publikation beschrieben und abgebildet. Auf methodische Schwächen/Stärken der eingeschlossenen Publikationen, sowie auf Aspekte, die die Übertragbarkeit der Ergebnisse beeinflussen, wurde im Text eingegangen.

Ergebnisse der Qualitätsbewertung:

Tabelle 41: Dokumentationsformular: Qualitätsbewertung

Quelle	Verwendeter Stufenplan		Punkte	Ein-/Ausschluss		Kommentar:
	1	2		ein	aus	
Homik & Hailey 1998	<input checked="" type="checkbox"/>		8 / 9	<input checked="" type="checkbox"/>		Qualitätsprüfung der Primärmaterialien erwähnt, aber nicht beschrieben.
Prins et al. 1997	<input checked="" type="checkbox"/>	<input type="checkbox"/>	entfällt	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Methodik fehlt, Abbruch der Bewertung nach Frage 2
Dretakis et al. 1998	<input type="checkbox"/>	<input checked="" type="checkbox"/>	entfällt	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Querschnittsauswertung einer prospektiv geplanten Studie
Garnero et al. 1998	<input type="checkbox"/>	<input checked="" type="checkbox"/>	entfällt	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Nested Case Control Study
Dargent-Molina et al. 1999	<input type="checkbox"/>	<input checked="" type="checkbox"/>	8	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Gnudi et al. 2000	<input type="checkbox"/>	<input checked="" type="checkbox"/>	8	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Huang et al. 1998	<input type="checkbox"/>	<input checked="" type="checkbox"/>	7 / 8	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Vollständigkeit des Follow-Up unklar
Pluijm et al. 1999	<input type="checkbox"/>	<input checked="" type="checkbox"/>	8	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Stewart et al. 1999	<input type="checkbox"/>	<input checked="" type="checkbox"/>	8	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

Datenextraktion: Dokumentation im Text, Tabelle

Schlussfolgerungen: Zuordnung der Evidenzlevel nach Hierarchie von CEBM (2001) (vergl. Tabelle 36)

Tabelle 42: Evidenzbewertung der eingeschlossenen Materialien

Quelle	Evidenzlevel nach CEBM, 2001
Homik&Hailey 1998	4*
Dargent-Molina et al. 1999	1b
Gnudi et al. 2000	1b
Huang et al. 1998	1b
Pluijm et al. 1999	1b
Stewart et al. 1999	1b

**Der HTA-Bericht erhielt die niedrige Einstufung, weil ihm eine Dokumentation der Qualität der eingeschlossenen Primärmaterialien fehlte. Der Evidenzlevel 4 beschreibt somit die "worst case" Annahme (dass es sich bei den eingeschlossenen Primärstudien um methodisch mangelhafte Arbeiten handelt).*