

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 35

Dünnschichtpräparationen und computergestützte Untersuchungen von Zervixabstrichen

Medizinische Effektivität, gesundheitsökonomische Evaluation und systematische Entscheidungsanalyse

Herausgeber:

**Uwe Siebert, MPH, MSc^{1,2,3}, Christiane Muth, MPH⁴, Gaby Sroczynski^{1,2},
Marcial Velasco-Garrido⁵, Dr. med. Ansgar Gerhardus, M.A.⁴,
Dr. med. Bernhard Gibis, MPH⁶, (Hrsg.)**

¹ Bayerische Forschungs- und Koordinierungsstelle Public Health, Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Klinikum Großhadern, Ludwig-Maximilians-Universität, München

² Harvard Center for Risk Analysis, Harvard School of Public Health, Boston, USA

³ Institute for Technology Assessment, Massachusetts General Hospital, Harvard Medical School, Boston, USA

⁴ Abteilung Epidemiologie, Sozialmedizin und Gesundheitssystemforschung, Medizinische Hochschule Hannover, Hannover

⁵ Institut für Gesundheitswissenschaften, FG Management im Gesundheitswesen, Technische Universität Berlin, Berlin

⁶ Kassenärztliche Bundesvereinigung (KBV), Dezernat 2, Versorgungsqualität und Sicherstellung, Berlin



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 270035

ISBN 3-537-27035-6

Druck: ICS Kommunikations-Service GmbH, Bergisch Gladbach

Gedruckt auf säurefreiem, alterungsbeständigem und chlorfreiem Papier.

Inhaltsverzeichnis

A Abstracts	1
A.1 Abstracts - English	1
I Abstract Clinical Effectiveness	1
II Abstract Economic Evaluation	3
III Abstract Systematic Decision Analysis.....	5
A.2 Abstracts - Deutsch	7
1 Abstract medizinische Effektivität	7
2 Abstract ökonomische Evaluation	9
3 Abstract systematische Entscheidungsanalyse.....	11
B Executive Summary.....	13
B.1 Medizinische Effektivität	13
B.2 Ökonomische Evaluation	17
B.3 Systematische Entscheidungsanalyse	25
I Medizinische Effektivität.....	37
I C Hauptdokument Medizinische Effektivität	39
I C 1 Policy Question	39
I C 2 Hintergrund.....	42
I C 2. 1 Beschreibung des Krankheitsbildes	42
I C 2. 2 Zervixkarzinom-Screening	62
I C 2. 3 Beschreibung der Technologie.....	74
I C 2. 4 Expertenmeinungen und Empfehlungen	83
I C 2. 5 Lebenszyklus der Technologie	84
I C 3 Forschungsfragen	86
I C 4 Methodik.....	87
I C 4. 1 Zielpopulation	87
I C 4. 2 Ergebnisparameter	87
I C 4. 3 Informationsquellen und Recherchen.....	92
I C 4. 4 Ein- und Ausschlusskriterien	93
I C 4. 5 Bewertung und Auswertung der Studien	95
I C 4. 6 Datenextraktion und Synthese	97
I C 5 Ergebnisse.....	99
I C 5. 1 Ergebnisse der Literaturrecherche	99
I C 5. 2 HTA-Berichte, systematischer Review und Metaanalyse.....	100
I C 5. 3 Primärstudien.....	133
I C 5. 4 Zusammenfassung der Ergebnisse – Effektivität	152

I C 6 Diskussion	156
I C 6. 1 Bezugsrahmen.....	156
I C 6. 2 Diskussion der Methodik.....	158
I C 6. 3 Inhaltliche Aspekte.....	166
I C 6. 4 Weiterer Forschungsbedarf	170
I C 6. 5 Zusammenfassung der Diskussion - Effektivität	172
I C 7 Schlussfolgerung	173
I C 8 Literatur	174
I C 8. 1 Literatur allgemein	174
I C 8. 2 Eingeschlossene Literatur	183
I C 8. 3 Ausgeschlossene Literatur	185
I Anhang	187
I 1 Vorsorgeschein	187
I 2 Dokumentation der Recherchen	188
I 3 Checklisten: HTA-Berichte und Systematische Übersichtsarbeiten	190
II Ökonomische Evaluation.....	197
II C Hauptdokument Ökonomische Evaluation	199
II C 1 Policy Question	199
II C 2 Hintergrund.....	199
II C 3 Forschungsfragen	200
II C 4 Methodik.....	201
II C 4. 1 Quellen und Recherchen	201
II C 4. 2 Expertenkontakte	205
II C 4. 3 Systematische Literaturselektion	207
II C 4. 4 Bewertung der Literaturstellen.....	211
II C 4. 5 Datenextraktion und qualitative Synthese	212
II C 4. 6 Währungskonversion und Inflationsbereinigung	213
II C 4. 7 Erhebung ökonomischer Primärdaten für Deutschland.....	213
II C 5 Ergebnisse.....	217
II C 5. 1 Ergebnisse der systematischen Literatursichtung.....	217
II C 5. 2 Kosten der Zervixkarzinom-Früherkennung und Therapie in Deutschland.....	224
II C 5. 3 Kosten der neuen Technologien.....	226
II C 5. 4 Ökonomische Konsequenzen des Einsatzes neuer Technologien im Zervixkarzinom-Früherkennungsprogramm.....	230
II C 6 Diskussion	259
II C 6. 1 Vergleichbarkeit und Limitationen	259
II C 6. 2 Kosten-Effektivität der neuen Screeningtechnologien.....	263
II C 6. 3 Übertragbarkeit der Ergebnisse auf den Kontext des deutschen Versorgungssystems.....	265
II C 7 Schlussfolgerungen.....	267
II C 8 Literatur	268
II Anhang	269
II 1 Dokumente der Literatursuche.....	269
II 2 Checkliste zur Beurteilung der methodischen Qualität gesundheitsökonomischer Studien	287

II 3 Studienmerkmale und Ergebnisse der eingeschlossenen gesundheitsökonomischen Studien	310
II 4 Einzelergebnisse der inhaltlich-qualitativen Bewertung evaluierter ökonomischer Studien.....	316
III Systematische Entscheidungsanalyse	321
III C Hauptdokument Systematische Entscheidungsanalyse	323
III C 1 Policy Question	323
III C 2 Hintergrund.....	323
III C 3 Forschungsfragen	324
III C 4 Methodik	325
III C 4. 1 Methodik der Markov-Modellierung	325
III C 4. 2 Evaluierte Strategien	326
III C 4. 3 Zielgrößen der Entscheidungsanalyse	327
III C 4. 4 Zeithorizont	328
III C 4. 5 Perspektive	328
III C 4. 6 Diskontierung.....	329
III C 4. 7 Gesellschaftliche Zahlungsbereitschaft.....	329
III C 4. 8 Beschreibung der Modellstruktur	329
III C 4. 9 Parameter des Entscheidungsmodells	332
III C 4. 10 Validierung des Modells	350
III C 4. 11 Modellannahmen	353
III C 4. 12 Szenarioanalysen, Subgruppenanalysen und Sensitivitätsanalysen	355
III C 4. 13 Software.....	357
III C 5 Ergebnisse	358
III C 5. 1 Effektivität	358
III C 5. 2 Kosten und Kosten-Effektivität	362
III C 5. 3 Szenarioanalysen	369
III C 5. 4 Subgruppenanalyse.....	371
III C 5. 5 Sensitivitätsanalysen	371
III C 6 Diskussion	375
III C 6. 1 Screeningintervall	376
III C 6. 2 Limitationen und weiterer Forschungsbedarf	378
III C 7 Schlussfolgerungen.....	384
III C 8 Literatur	386
III C 8. 1 Allgemeine Literatur aus den Kapiteln II und III.....	386
III C 8. 2 Eingeschlossene Studien ökonomische Evaluation (Kapitel II)	413
III C 8. 3 Ausgeschlossene Studien ökonomische Evaluation (Kapitel II)	414
III C 8. 4 Im entscheidungsanalytischen Modell berücksichtigte Literatur (Kapitel III).....	415
Glossar	421

Tabellenverzeichnis

Tabelle 1:	Klassifikation zytologischer Befunde in der Münchner Nomenklatur II (Schenck 2001).....	51
Tabelle 2:	Epitheldysplasien im Bethesda-System 1991 (Ausschnitt).....	52
Tabelle 3:	Schematische Gegenüberstellung von Nomenklaturen.....	53
Tabelle 4 :	Stadieneinteilung des Zervixkarzinoms (modifiziert nach Riethdorf et al. 1999b)	55
Tabelle 5:	Empfehlungen zur operativen Therapie bei Zervixdysplasien	58
Tabelle 6:	Stadienadaptierte Therapie des Zervixkarzinoms in Deutschland.....	59
Tabelle 7:	Standardoperationsverfahren bei invasivem Zervixkarzinom	59
Tabelle 8:	Nachsorgeprogramm für klinisch tumorfreie Patienten in Deutschland.....	61
Tabelle 9:	Einsatzbereiche von Geräten und Instrumenten im Screeningablauf.....	75
Tabelle 10:	Erwartete Vor- und Nachteile der neuen Technologien im Vergleich zum herkömmlichen Pap Test.....	81
Tabelle 11:	Phaseneinteilung diagnostischer Studien	88
Tabelle 12:	Studiendesigns diagnostischer Studien in Abhängigkeit zum Untersuchungsgegenstand.....	88
Tabelle 13:	Testgütekriterien.....	90
Tabelle 14:	Prädiktionswerte.....	92
Tabelle 15:	HTA-Berichte anderer Organisationen zur Technologiebewertung – gesamt	101
Tabelle 16:	Zusammenfassung der berücksichtigten HTA-Reports zu klinischer Effektivität.....	126
Tabelle 17:	Systematischer Review / Metaanalyse zu Dünnschichtpräparationstechniken	130
Tabelle 18:	Übersicht über Qualitätsscores der berücksichtigten Publikationen.....	133
Tabelle 19:	Evidenztabelle der Studien über computergesteuerte Technologien zur Auswertung von Zervikalabstrichen	138
Tabelle 20:	Evidenztabelle der Studien über Dünnschichtpräparationen von Zervikalabstrichen	147
Tabelle 21:	Dünnschichtpräparationstechniken im Zervixkarzinom-Screening: Ausgeschlossene Studien	151
Tabelle 22:	Zusammenfassung der berücksichtigten Sekundärpublikationen	153
Tabelle 23:	Zusammenfassung der berücksichtigten Primärpublikationen	154
Tabelle 24:	Testgütekriterien mit/ohne Einschluss der inadäquaten Präparate	166
Tabelle 25:	Publikationsliste internationaler Einrichtungen	203
Tabelle 26:	Internetseiten der berücksichtigten Institutionen.....	204

Tabelle 27:	Ein- und Ausschlusskriterien der Literatursuche. Gruppe A: Verfahren zur Früherkennung des Zervixkarzinoms	208
Tabelle 28:	Ein- und Ausschlusskriterien der Literatursuche. Gruppe B: Gesundheitsökonomie	209
Tabelle 29:	Ein- und Ausschlusskriterien der Literatursuche. Gruppe C: Zervixkarzinom	210
Tabelle 30:	Ein- und Ausschlusskriterien der Literatursuche. Gruppe D: Entscheidungsanalytische Modelle	211
Tabelle 31:	Anzahl, Publikationsart und Gruppenzugehörigkeit eingeschlossener Literatur.....	217
Tabelle 32:	Ausgeschlossene ökonomische Studien mit Ausschlussgrund	219
Tabelle 33:	Manuelle Selektion zur qualitativen Bewertung ökonomischer Studien (Gruppe B) ..	220
Tabelle 34:	Prozentuale Erfüllung der Qualitätskriterien und Summenscore evaluierter ökonomischer Studien	223
Tabelle 35:	Mittlere Gesamtkosten für Früherkennungsuntersuchung, Therapie, Medikation und Nachsorge des Zervixkarzinoms in Deutschland (gerundete Werte).....	226
Tabelle 36:	Anbieter und Preisangaben neuer Technologien.....	227
Tabelle 37:	Kosten der Dünnschichtpräparationen nach Payne et al. 2000 (Payne et al. 2000) ..	227
Tabelle 38:	Kostenberechnung für die ThinPrep™-Technologie (Kosten/Präparat)	228
Tabelle 39:	Kostenberechnung für die AutoCytePrep™-Technologie (Kosten/Präparat)	229
Tabelle 40:	Kostenberechnung für die AutoPap™ GS Screeningtechnologie (Kosten/Präparat)	229
Tabelle 41:	Internationale gesundheitsökonomische Studien	231
Tabelle 42:	Übersicht der in den evaluierten gesundheitsökonomischen Studien berücksichtigten Kostenarten	249
Tabelle 43:	Inkrementelle Kosten und Effekte von neuen Primärscreeningverfahren im Vergleich zum konventionellen Pap-Test	252
Tabelle 44:	Inkrementelle Kosten und Effekte von neuen Rescreeningverfahren im Vergleich zum konventionellen Pap-Test	253
Tabelle 45:	Sensitivitäten und Spezifitäten konventioneller und neuer Primärscreening-Verfahren	255
Tabelle 46:	Sensitivität und Spezifität manueller und automatisierter Rescreening-Verfahren.....	257
Tabelle 47:	Studienmerkmale und Ergebnisse der eingeschlossenen gesundheitsökonomischen Studien.....	310
Tabelle 48:	Inhaltlich-qualitative Bewertung der evaluierten ökonomischer Studien, Einzelergebnisse.	316
Tabelle 49:	Jährliche Übergangswahrscheinlichkeiten im Markov-Modell	334
Tabelle 50:	Verteilung der Zervixkarzinomstadien nach FIGO in Prozent. Quelle: Daten der Registerstelle des Gemeinsamen Krebsregisters (GKR) Berlin (Jahr 1964-66).....	337

Tabelle 51:	Altersspezifische jährliche benigne Hysterektomie rate aufgrund anderer Ursachen als ein Zervixkarzinom. Quelle: (Vessey & Villard-Mackintosh 1992), UK (Daten von 1985-1989).....	338
Tabelle 52:	Jährliche Überlebenswahrscheinlichkeiten nach Zervixkarzinomstadien (FIGO) in Prozent. Quelle: Tumorregister München, Daten Region München von 1988-1997.....	339
Tabelle 53:	Sensitivität und Spezifität des Papanicolaou-Screeningtests (nach Nanda et al. 2000) und Anteile der milden und moderat-schwerwiegenden Testergebnisse (nach Payne et al. 2000) für die Stadien CIN 1, CIN 2, CIN 3 und invasives Karzinom.....	342
Tabelle 54:	Sensitivität der Dünnschichtpräparation. Schwellenwert LSIL+/CIN 1+.....	344
Tabelle 55:	Sensitivität der Dünnschichtpräparation. Schwellenwert LSIL+/CIN 2+.....	345
Tabelle 56:	Sensitivität der automatisierten Primärscreeningverfahren. Schwellenwert LSIL+/CIN 1+ und Schwellenwert LSIL+/CIN 2+.....	346
Tabelle 57:	Sensitivität der Dünnschichtpräparation mit anschließendem automatisierten Primärscreeningverfahren. Schwellenwert LSIL+/CIN 1+.....	347
Tabelle 58:	Undiskontierte Kosten für Screening, Diagnostik, Behandlung und Palliativmaßnahmen des Zervixkarzinoms.....	349
Tabelle 59:	Vergleich der Modellprädiktion mit Beobachtungsdaten für eine ungescreente deutsche Population (interne und externe Modellvalidierung).....	352
Tabelle 60:	Inkrementelle undiskontierte Lebenserwartung für die untersuchten Strategien bei verschiedenen Screeningintervallen.	360
Tabelle 61:	Absolute und inkrementelle verhinderte Krebsfälle pro 100.000 Frauen für die untersuchten Strategien bei verschiedenen Screeningintervallen.	361
Tabelle 62:	Absolute und inkrementelle verhinderte Krebstodesfälle pro 100.000 Frauen für die untersuchten Strategien bei verschiedenen Screeningintervallen.	362
Tabelle 63:	Basisfall-Analyse: Absolute Kosten und Effektivität sowie inkrementelle diskontierte Kosten-Effektivitäts-Verhältnisse für die untersuchten Strategien bei einem jährlichen Screeningintervall (gerundete Werte).	364
Tabelle 64:	Szenarien für verschiedene Screeningintervalle: Absolute undiskontierte Kosten und Effektivität sowie diskontierte inkrementelle Kosten-Effektivitäts-Verhältnisse für die untersuchten Strategien bei verschiedenen Screeningintervallen.	367

Abbildungsverzeichnis

Abbildung 1:	Screeningteilnahmeraten, Inzidenz und Mortalität des Zervixkarzinoms in den USA	43
Abbildung 2:	Altersspezifische Inzidenz des Zervixkarzinoms im Saarland	44
Abbildung 3:	Zeitlicher Verlauf der Zervixkarzinominzidenz in Deutschland	45
Abbildung 4:	Altersstandardisierte Zervixkarzinom-Mortalitätsrate in Deutschland	46
Abbildung 5:	Karzinogenese der Zervix uteri	49
Abbildung 6:	Stadienspezifische Überlebensraten beim invasiven Zervixkarzinom	62
Abbildung 7:	Lebenszyklus einer Technologie	85
Abbildung 8:	Diagramm für die Darstellung einer ROC-Kurve	91
Abbildung 9:	Einfluss der inkrementellen Sensitivität auf das inkrementelle Kosten-Effektivitäts-Verhältnis des Einsatzes neuer Primärscreening-Technologien versus konventioneller Methode.	256
Abbildung 10:	Einfluss der inkrementellen Sensitivität auf das inkrementelle Kosten-Effektivitäts-Verhältnis des Einsatzes automatisierter Rescreening-Technologien versus konventioneller Methode.	258
Abbildung 11:	Markov-Modell in Form eines Bubble Diagram	332
Abbildung 12:	Absolute Lebenszeitkosten der untersuchten Screeningverfahren (in €) und inkrementelle gewonnene Lebensstage versus keinem Screening	365
Abbildung 13:	Szenario: Kosten-Effektivitäts-Frontier für die untersuchten Strategien bei verschiedenen Screeningintervallen (Diskontierte Kosten und Effekte)	369
Abbildung 14:	Einweg-Sensitivitätsanalyse: Inkrementelle Kosten-Effektivitäts-Verhältnisse (in €/Lebensjahr) der untersuchten Strategien bei unterschiedlicher unterer Screening-Altersgrenze	370
Abbildung 15:	Einweg-Sensitivitätsanalyse: Inkrementelle Kosten-Effektivitäts-Verhältnisse (in €/Lebensjahr) der untersuchten Strategien bei unterschiedlichen Teilnahmeraten am Screening.	373
Abbildung 16:	Zwei-Weg-Sensitivitätsanalyse: Minimum, Basisfall und Maximum der inkrementellen Kosten-Effektivitäts-Verhältnisse (in €/Lebensjahr) der untersuchten Strategien bei Variation von Kosten und Sensitivitäten bei einjährigem Screeningintervall.	374

Abkürzungsverzeichnis

ASCUS	Atypical Squamous Cell of Uncertain Significance
AGUS	Atypical Glandular Cells of Undetermined Significance
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
BMG	Bundesministerium für Gesundheit
CIN	Zervikale intraepitheliale Neoplasie
CIS	Carcinoma in situ
CLIA	Clinical Laboratory Improvement Act
DAHTA	Deutsche Agentur für Health Technology Assessment
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information
EBM	Einheitlicher Bewertungsmaßstab
FDA	Food and Drug Administration
FIGO	Fédération Internationale de Gynécologie et d'Obstétrique
FN	False negative
FNR	False Negative Rate
FP	False positive
FPR	False positive rate
GKV	Gesetzliche Krankenversicherung
GOÄ	Gebührenverordnung für Ärzte
GRV	Gesetzliche Rentenversicherung
HPV	Humanes Papillomavirus
HSIL	High Grade Squamous Intraepithelial Lesion
HTA	Health Technology Assessment
IARC	International Agency for Research on Cancer

INAHTA	International Network of Agencies for Health Technology Assessment
ISTAHC	International Society of Technology Assessment in Health Care
IKEV	Inkrementelles Kosten-Effektivitäts-Verhältnis
LEEP	Loop electrosurgical excision procedure
LLETZ	Large loop excision of the transformation zone
LSIL	Low Grade Squamous Intraepithelial Lesion
NPV	Negative predictive value
OR	Odds Ratio
PAP	Papanicolaou
PPV	Positive predictive value
RCT	Randomized clinical trial
ROC	Receiver operating characteristic
SEER	National Cancer Institute's Surveillance, Epidemiology, and End Results
SIL	Squamous intraepithelial lesion
SROC	Summary receiver operating characteristic
TN	True negative
TNM	Internationale Einteilung der Krankheitsstadien bei Krebserkrankungen (Tumor, Nodus, Metastasen)
TP	True positive
TPR	True positive rate
UICC	Union Internationale Contre le Cancer
WHO	World Health Organisation

A Abstracts

A.1 Abstracts - English

I Abstract Clinical Effectiveness

C. Muth, G. Sroczynski, M. Velasco-Garrido

Background: The introduction of cervical cancer screening programmes has led to a worldwide reduction of cervical cancer incidence. In Germany, annual screening with the Papanicolaou smear (the “Pap test”) is granted by the German Social Code Book for all women over the age of twenty years. Since the introduction of the screening test in the 1970s, a decay in the incidence and mortality from cervical cancer has been observed in Germany. However some concerns have been raised about the sensitivity of the test. The false negative rate can be reduced with quality assurance measures. The use of liquid-based preparations and automated devices could lead to further reductions in the false negative rate. In Germany, such technologies are at an early stage of diffusion.

Objectives:

- To assess the sensitivity and specificity of liquid-based slide preparations and of automated devices as compared to conventional Papanicolaou smear.
- To assess the clinical effectiveness of the new technologies as measured by reductions of incidence and mortality from cervical cancer.

Methods: A systematic literature search was conducted in February 2001 and updated in November 2001 of the following databases: Medline, EMBASE, CancerLit, Science Citation Index, the Cochrane Library, and the HTA-Database. The literature search was supplemented with an Internet-based hand search of HTA Institutions, Professional Associations, and Guidelines Clearing Houses. The quality of HTA reports and systematic reviews was assessed using checklists developed by the

German Scientific Working Group on Technology Assessment. The quality of primary studies was assessed using an adapted version of the score from Flynn and Adams (1996). To be included in the analysis, primary studies must have reported at least one of the following outcomes: sensitivity, specificity, changes in cervical cancer incidence or mortality.

Information was synthesised in evidence tables and discussed. Quantitative synthesis was not possible because of the great heterogeneity in published studies.

Results: A total of four HTA reports, a meta-analysis and six primary studies were included in this report. None of the documents assessed changes in incidence or mortality from cervical cancer. The HTA reports agreed that the current evidence does not allow conclusions concerning the sensitivity or specificity of the new technologies, since quality of primary studies is poor and bias can not be excluded. The quality of the six primary studies included in our report was also limited mainly due to a lack of adequate reference standards. Reported estimates on sensitivity or specificity are thus lacking in validity for both kinds of technologies. For liquid-based preparation techniques, the evidence indicates that they may improve the quality of smears. However, the relevance of this finding for the effectiveness of the screening is still under discussion.

Conclusions: There is not enough evidence that the use of new technologies such as liquid-based preparations and automated devices leads to higher effectiveness than the conventional Papanicolaou smear when screening for cervical cancer. Liquid-based preparation techniques can probably improve the quality of smears, however this could also be achieved by a more carefully conducted smear collection. Further studies are needed to evaluate the sensitivity and specificity of the new technologies and their effect on morbidity and mortality from cervical cancer. There is a need for evaluation of different strategies for helping more women take advantage of the German screening programme, since higher participation quotes could contribute to further reduction of cervical cancer mortality rather than improvement of conventional testing.

II Abstract Economic Evaluation

U. Siebert, G. Sroczynski, K. Voigt, P. Aidelsburger, J. Engel, B. Gibis, P. Hillemanns, D. Hölzel, J. Wasem

Introduction: This health technology assessment (HTA) commissioned by the German Agency for HTA at the German Institute for Medical Documentation and Information (DAHTA@DIMDI) included a systematic review of the evidence on the cost-effectiveness of new cervical cancer screening techniques (liquid-based cytology and automated screening procedures). A further objective was to assess the transferability of the current evidence to the context of the German health care system, and to identify areas in which future research is needed.

Methods: A systematic literature review and selection of appropriate studies were performed. Study quality and transferability of evidence to the German health care context were assessed using instruments developed by the German Scientific Working Group Technology Assessment for Health Care. The economic evidence was systematically synthesized and discussed.

Results: Based on international studies, the discounted incremental cost-effectiveness ratios (ICER) compared to conventional smears were 10,000–37,000 US\$ per life year saved (LYS) for liquid based cytology, 22,000–29,000 US\$/LYS for automated primary screening, and 16,000 – 48,000 US\$/LYS for automated re-screening technologies, assuming a 3-year screening interval. Shorter screening intervals resulted in higher ICERs. Assuming a societal willingness-to-pay of 50,000 US\$/LYS, new cervical cancer screening techniques may be cost-effective at a screening interval of 3 years or longer.

The transferability of these results to the German health care context is questionable. Demography, epidemiology of the disease, clinical practice patterns, and clinical guidelines, as well as the structure of the health care system, prices, and preferences in Germany differ substantially from those in other countries. An adaption may be difficult to perform, especially in the absence of detailed data. An alternative is the development of a decision-analytic model specific for the Germany health care context.

Conclusions: Based on international studies, new technologies aimed at improving the performance of cervical cancer screening are reasonably cost-effective at screening intervals of 3 years or longer. Based on the current clinical standard of annual cervical cancer screening in Germany, the introduction of the evaluated new scree-

ning technologies would likely not be cost-effective. A decision-analytic cost-effectiveness study on German screening and treatment practice is necessary to yield valid economic informations.

Further research should be performed to precisely assess the costs of new screening techniques, as well as to provide valid estimates for sensitivity and specificity of conventional and new screening techniques in a routine setting, in a representative population, and using a gold standard. In addition, research should be performed on the effect of improving screening compliance, especially in underprivileged subpopulations with low screening compliance, and the effect of changing the German screening practice to longer screening intervals.

III Abstract Systematic Decision Analysis

U. Siebert, G. Sroczynski, B. Gibis, P. Aidelsburger, J. Engel, P. Hillemanns, K. Voigt, C. Muth, M. Velasco-Garrido, J. Wasem, D. Hölzel, S.J. Goldie

Introduction: The objective of this decision analysis was to develop a decision analytic model for the German health care context to evaluate the clinical and economic long-term consequences of new cervical cancer screening techniques considering German clinical and screening practice, and to discuss the uncertainty regarding the current evidence.

Methods: A decision-analytic Markov model was developed to evaluate the clinical and economic long-term consequences of different cervical cancer screening strategies. German epidemiologic and economic data were used and information about German clinical practice in screening, diagnosis, and treatment of cervical cancer and its precursors were considered. Evaluated outcomes included detected cervical cancer cases, cervical cancer deaths, life expectancy, lifetime costs, and cost-effectiveness of the compared screening strategies.

Results: Compared to no screening, annual conventional screening resulted in a discounted incremental cost-effectiveness ratio (ICER) of 6,000 € per life-year saved (LYS) for the German health care context. Compared to annual conventional screening, the ICER was 220,000 €/LYS for automated smear analysis, 953,000 €/LYS for liquid-based preparation, and 1,083,000 €/LYS for a combination of liquid-based preparation and automated smear analysis. Less frequent screening resulted in lower ICERs. The ICER for automated smear analysis compared to conventional screening decreased to 26,000 €/LYS for a 2-year screening interval. Similarly, for a 3-year screening interval, the ICER was 46,000 €/LYS for liquid-based preparation and 53,000 €/LYS for a combination of liquid-based preparation and automated smear analysis. Assuming a societal willingness-to-pay of 50,000 €/LYS, new cervical cancer screening techniques were cost-effective at screening intervals equal to or greater than 2 or 3 years, respectively. Screening every two years with new cervical cancer screening techniques may be as effective as annual conventional screening, but may generate lower costs.

Conclusion: Based on our analysis for the German health care system, new technologies aimed at improving the performance of cervical cancer screening are reasonably cost-effective at screening intervals of 2 years and longer. For the current

clinical standard of annual cervical cancer screening in Germany, the introduction of new screening technologies would likely not be cost-effective without a reduction in screening frequency.

A reduction in screening frequency to every other year should be critically discussed within the context of improving screening compliance of women. Further research should focus on providing valid estimates for sensitivity and specificity of conventional and new screening techniques in the routine setting, in a representative population, and using a gold standard. In addition, research should be performed to improve the screening compliance of women, especially in the context of changing German screening practice to less frequent cervical cancer screening practice.

A.2 Abstracts - Deutsch

1 Abstract medizinische Effektivität

C. Muth, G. Sroczynski, M. Velasco-Garrido

Hintergrund: Vor Einführung von Maßnahmen zur Früherkennung zählte der Gebärmutterhalskrebs zu den häufigsten bösartigen Erkrankungen der Frau. Seit Einführung des Papanicolaou Abstrichtests (Pap Test) als Zervixkarzinom-Screening in das Früherkennungsprogramm - nach §25 Abs. 2 SGB V haben alle Frauen ab dem 20. LJ. einmal jährlich Anspruch auf diese Untersuchung - fielen Inzidenz und Mortalität an diesem Tumor deutlich ab. Der Pap Test stellt jedoch angesichts seiner Testgütekriterien keinen idealen Screeningtest dar, da die Rate falsch negativer Befunde, bezogen auf den Einzeltest, beträchtlich sein kann. Diese Rate lässt sich durch Qualitätskontrollen verbessern. Darüber hinaus sind neue Technologien entwickelt worden, welche die diagnostische Genauigkeit und Zuverlässigkeit von Abstrichuntersuchungen im Screeningprozess verbessern sollen und die sich in Deutschland derzeit in der Phase der Markteinführung befinden.

Fragestellung: Lässt sich in diagnostischen Studien die Testgüte der Dünnschichtpräparationen und der computergestützten Auswertungen bestimmen und wie unterscheiden sich die genannten neuen Technologien darin vom konventionellen Pap Test? Gibt es darüber hinaus Erkenntnisse in der Literatur, die eine prognostische Aussage zulassen, inwieweit die neuen Verfahren Inzidenz und Mortalität des Zervixkarzinoms positiv beeinflussen?

Methodik: Zielpopulation sind alle Frauen ab dem 20. LJ. Als Ergebnisparameter wurden Testgütekriterien, wie z.B. Sensitivität und Spezifität gewählt und für prognostische Studien Angaben zur Senkung von Inzidenz und Mortalität. Im Februar 2001 wurde eine systematische Literaturrecherche in den elektronischen Datenbanken Medline, EMBASE, CancerLit und Science Citation Index, in den Datenbanken internationaler HTA-Institutionen, kombiniert mit einer Handsuche auf Homepages von Leitlinienherstellern und Fachgesellschaften und in den Publikationslisten bereits ermittelter Dokumente durchgeführt, ein Recherche-Update erfolgte im November 2001. Die identifizierten Dokumente mussten definierte Kriterien erfüllen. Die methodische Qualität von HTA-Berichten und Übersichtsarbeiten wurden anhand von

Checklisten bewertet, primäre Studien wurden bezüglich ihrer methodischen Qualität in eine von vier Qualitätsstufen klassifiziert. Die Synthese der Daten erfolgte narrativ, eine quantitative Synthese war aufgrund großer Heterogenität der Studien nicht sinnvoll.

Ergebnisse: Insgesamt wurden vier HTA-Berichte und eine Metaanalyse sowie sechs primäre Studien zu diagnostischer Genauigkeit und Zuverlässigkeit in die Evaluation eingeschlossen, wobei keine in Deutschland durchgeführt worden war. Zu prognostischen Fragestellungen wurde keine Studie identifiziert.

Übereinstimmend wurde in allen HTA-Berichten eine unzureichende Evidenzlage für die Testgütekriterien bei Dünnschichtpräparationen wie auch bei computergestützter Auswertung von Pap Tests konstatiert und es wurde über eine überwiegend mangelhafte Qualität bei den durchgeführten klinischen Studien berichtet. Die von uns eingeschlossenen Primärpublikationen wiesen überwiegend inadäquate Referenzstandards und/oder schwerwiegende methodische Mängel auf. Diese Qualitätsmängel begründeten die unzureichende interne und externe Validität der Ergebnisse. Weder für Dünnschichtpräparationen, noch für computergestützte Auswertungen von Abstrichen konnten von uns zuverlässige Angaben zur Testgüte ermittelt werden. Auf der Stufe technischer Machbarkeitsprüfungen von diagnostischen Tests ergaben sich Hinweise, dass Dünnschichtpräparationstechniken einen höheren Anteil adäquater Abstriche liefern können, diese Hinweise waren jedoch nicht konsistent.

Schlussfolgerung: Aus den hier ausgewerteten Primärstudien und Kontextdokumenten konnte keine Evidenz für eine Überlegenheit von computergestützten Auswertungen oder von Dünnschichtpräparationen gegenüber konventionellen Pap Tests gewonnen werden, da die diagnostische Genauigkeit und Zuverlässigkeit der neuen Technologien noch nicht hinreichend genau bestimmt worden sind. Auf der Stufe technischer Machbarkeitsprüfungen liegt für die Dünnschichtpräparationstechniken eine unsichere Evidenzlage vor, ob sie in der Lage sind, eine größere Zahl adäquater Proben zu liefern oder ob bei Beachtung wesentlicher Grundregeln der Sorgfalt der konventionelle Pap Test ebenso gute Ergebnisse liefert. Auch ist in der Literatur die Diskussion noch nicht abgeschlossen, dass Dünnschichtpräparationen gleichzeitig einen Informationsverlust und eine bessere Auswertbarkeit der Präparate bedeuten. Es besteht somit weiterer Forschungsbedarf, um präzise und valide Ergebnisse zu Testgenauigkeit und Zuverlässigkeit beider Technologien einzeln sowie in einer Kombination aus beiden Verfahren zu ermitteln. Darüber hinaus besteht Forschungsbedarf zu der Frage, welche Interventionen die Teilnahmeraten am Screening wirkungsvoll steigern können, da damit größere Effekte auf das Gesamt-screeningprogramm zu erwarten sind, als durch die Steigerung von Sensitivität und Spezifität des einzelnen Tests.

2 Abstract ökonomische Evaluation

U. Siebert, G. Sroczynski, K. Voigt, P. Aidelsburger, J. Engel, B. Gibis, P. Hillemanns, D. Hölzel, J. Wasem

Hintergrund: Im Rahmen dieses Health Technology Assessments (HTA) im Auftrag der Deutschen Agentur für Health Technology Assessment des Deutschen Instituts für Medizinische Information und Dokumentation (DAHTA@DIMDI) wurde ein systematischer Review zur internationalen Evidenz bezüglich der Kosten-Effektivität neuer Technologien des Zervixkarzinom-Screenings (Dünnschichtpräparation und computergestützte Auswertungsverfahren) durchgeführt. Es sollte ferner die Übertragbarkeit der Daten auf das deutsche Gesundheitssystem beurteilt und weiterer Forschungsbedarf identifiziert werden.

Methoden: Es wurde eine systematische Literaturrecherche und eine systematische Literaturselektion anhand a priori festgelegter Kriterien durchgeführt. Die Studienqualität wurde mittels standardisierter Checklisten der German Scientific Working Group Technology Assessment for Health Care bewertet und die Übertragbarkeit auf das deutsche Gesundheitssystem beurteilt. Die Ergebnisse wurden systematisch zusammengefasst und abschließend diskutiert.

Ergebnisse: In internationalen Studien betrug das diskontierte inkrementelle Kosten-Effektivitäts-Verhältnis (IKEV) bei einem 3-jährigen Screeningintervall für die Dünnschicht-Präparation 10.000-37.000 US\$ pro gewonnenem Lebensjahr (LJ), für automatisierte Primärscreening-Verfahren 22.000-29.000 US\$/LJ und für automatisierte Rescreening-Verfahren 16.000-48.000 US\$/LJ (jeweils im Vergleich zum konventionellen Papanicolaou-Verfahren). Kürzere Screeningintervalle führten zu höheren IKEV. Bei einer angenommenen Grenze der gesellschaftlichen Zahlungsbereitschaft von 50.000 US\$/LJ können die untersuchten Screening-Technologien ab einem 3-Jahres-Screening-Intervall kosteneffektiv eingesetzt werden.

Die Übertragbarkeit dieser Ergebnisse auf das deutsche Gesundheitssystem ist fraglich. Demographie, Epidemiologie der Erkrankung, klinische Praxis und klinische Richtlinien sowie Struktur des Gesundheits- und Vergütungssystems, Preise und Verbraucherpräferenzen in Deutschland unterscheiden sich zum Teil deutlich von denen anderer Länder. Die notwendigen Verfahren zur Anpassung der Ergebnisse an den deutschen Kontext sind sehr aufwendig und in vielen Bereichen aufgrund mangelnder Datenlage nicht möglich. Als Alternative ist die Entwicklung eines für Deutschland spezifischen entscheidungsanalytischen Modells zu sehen.

Schlussfolgerungen: Internationale Studien berichteten einen kosteneffektiven Einsatz von neuen Zervixkarzinom-Screeningverfahren bei einem Screeningintervall von 3 oder mehr Jahren. Für den derzeitigen deutschen Standard des jährlichen Screenings ist die Einführung der untersuchten neuen Screeningverfahren als nicht kosteneffektiv einzuschätzen. Um valide ökonomische Aussagen im Kontext der deutschen Screening- und Behandlungspraxis zu liefern, ist die Durchführung einer entscheidungsanalytischen Kosten-Effektivitätsstudie notwendig.

Forschungsbedarf besteht bei der präziseren Erfassung der Testkosten der neuen Verfahren sowie der Bestimmung der Testgüte konventioneller und neuer Screeningverfahren unter Routinebedingungen und unter Verwendung von Goldstandards bei der durchschnittlichen Bevölkerung. Ferner sollte der Effekt einer Verbesserung der Teilnahme der Frauen am regelmäßigen Screening, insbesondere in unterversorgten Subpopulationen, und der Effekt einer möglichen Verlängerung des Screeningintervalls systematisch analysiert werden.

3 Abstract systematische Entscheidungsanalyse

U. Siebert, G. Sroczynski, B. Gibis, P. Aidelsburger, J. Engel, P. Hillemanns, K. Voigt, C. Muth, M. Velasco-Garrido, J. Wasem, D. Hölzel, S.J. Goldie

Hintergrund: Ziel dieser systematischen Entscheidungsanalyse war, die medizinischen und ökonomischen Langzeit-Konsequenzen des Einsatzes neuer Screening-technologien in der Zervixkrebsfrüherkennung unter Einbeziehung derzeitiger wissenschaftlicher Evidenz und der für Deutschland spezifischen klinischen und gesetzlich festgelegten Praxis zu modellieren sowie die Relevanz der derzeitigen Informationsunsicherheit zu beurteilen.

Methoden: Es wurde ein entscheidungsanalytisches Markov-Modell zur klinischen und ökonomischen Evaluation von Langzeit-Konsequenzen verschiedener Strategien zum Zervixkarzinom-Screening entwickelt. Dabei wurden deutsche epidemiologische Daten, Angaben zur aktuellen deutschen Praxis in Diagnose und Behandlung des Zervixkarzinoms und seiner Vorstufen sowie deutsche Kostendaten verwendet. Zielgrößen waren entdeckte Karzinomfälle, zervixkarzinomspezifische Mortalität, die Lebenserwartung, die Lebenszeitkosten und die inkrementelle Kosten-Effektivität verschiedener Strategien für Deutschland.

Ergebnisse: Die Analysen mit dem deutschen Modell ergaben für ein jährliches konventionelles Screening im Vergleich zu keinem Screening ein inkrementelles Kosten-Effektivitäts-Verhältnis (IKEV) von 6.600 €/LJ. Im Vergleich zum konventionellen Screening lag das IKEV für die computergestützte Auswertung von Papanicolaou-Abstrichen bei 220.000 €/LJ, für die Dünnschicht-Präparation bei 953.000 €/LJ und für die Kombination von Dünnschicht-Präparation mit computergestützter Auswertung bei 1.083.000 €/LJ. Mit steigender Länge des Screeningintervalls sanken die IKEVs. Das IKEV des Screenings mit computergestützter Auswertung des Papanicolaou-Tests im Vergleich zum konventionellen Screening verringerte sich auf 26.000 €/LJ beim 2-Jahres-Intervall und das IKEV des Screenings mit Dünnschicht-Präparation mit manueller bzw. computergestützter Auswertung auf 46.000 bzw. 53.000 €/LJ beim 3-Jahres-Intervall, so dass diese bei einer angenommenen Grenze der gesellschaftlichen Zahlungsbereitschaft von 50.000 €/LJ bereits ab einem 2 bzw. 3-Jahres-Screening-Intervall als kosteneffektiv einzuschätzen sind. Bei einem zwei-jährigen Screeningintervall sind die neuen Technologien kaum weniger effektiv als das jährliche konventionelle Screening, aber verursachen zum Teil deutlich weniger Kosten.

Schlussfolgerungen: Basierend auf eigenen Analysen für das deutsche Gesundheitssystem können neue Zervixkarzinom-Screeningverfahren zu einer akzeptablen Kosten-Effektivität führen, wenn das Screeningintervall 2 Jahre oder mehr beträgt. Für den derzeitigen deutschen Standard des jährlichen Screenings wird die Einführung der neuen Screeningverfahren als nicht kosteneffektiv bewertet.

Unter der Gewährleistung einer ausreichenden Beteiligungsrate der Frauen am Screeningprogramm sollte jedoch kritisch in Betracht gezogen werden, ob ein Einsatz der neuen Screeningmethoden in Verbindung mit einer Verlängerung des Screeningintervalls auf 2-Jahre angestrebt werden sollte. Dabei sollten jedoch unbedingt die möglichen Auswirkungen eines solchen Vorgehens auf das Screening-Teilnahmeverhalten berücksichtigt werden. Forschungsbedarf besteht für die Testgüte konventioneller und neuer Screeningverfahren unter Routinebedingungen in Deutschland. Ferner zum Screening-Teilnahmeverhalten, insbesondere im Hinblick auf eine mögliche Verlängerung des Screeningintervalls.

B Executive Summary

B.1 Medizinische Effektivität

C. Muth, G. Sroczynski, M. Velasco-Garrido

Hintergrund

Vor Einführung von Maßnahmen zur Früherkennung zählte der Gebärmutterhalskrebs zu den häufigsten bösartigen Erkrankungen der Frau. Seit 1971 fiel mit Einführung des Zervixkarzinom-Screenings die altersstandardisierte Inzidenzrate über alle Altersgruppen von 40,5 / 100.000 auf 11,20 / 100.000 im Jahre 1996 (Robert-Koch-Institut; 1999). Im gleichen Zeitraum fiel auch die Mortalität deutlich ab, in den letzten 15 Jahren allein um 40% (Gesundheitsbericht für Deutschland; 1998). Diese positiven Entwicklungen werden u.a. auf ein effektives Screening zurückgeführt (Hillemanns et al.; 1997). In Deutschland haben Frauen nach §25 Abs. 2 SGB V ab dem 20. Lebensjahr einmal jährlich Anspruch auf eine Früherkennungsuntersuchung auf Zervixkarzinome. Diese erfolgt als Abstrichuntersuchung nach Papanicolaou, dem sogenannten Pap Test. Dabei handelt es sich um einen Abstrich, der speziell gefärbt im Labor nach auffälligen Zellen durchgemustert wird, die Dysplasien und damit potentiellen Krebsvorstufen oder bereits eindeutig malignen, also Krebszellen entsprechen. Ziel ist es, möglichst frühe Erkrankungsgrade zu erkennen, da Krebsvorstufen fast immer heilbar sind und bei bereits bestehenden invasiven Zervixkarzinomen das Überleben im wesentlichen vom Stadium der Erkrankung bei Diagnosestellung abhängt. Zur Erreichung dieses Ziels benötigt man Screeningtests, die eine möglichst hohe Sensitivität besitzen, also möglichst wenige Erkrankungen übersehen, die andererseits jedoch auch eine möglichst hohe Spezifität aufweisen, also möglichst wenige gesunde Frauen fälschlicherweise als krank einstufen. Der Pap Test zeigte in einer Metaanalyse von McCrory et al. (1999) eine überraschend geringe Sensitivität von 51 % angesichts der positiven Entwicklungen in Inzidenz und Mortalität seit seiner Einführung. Zur Senkung der Fehlerrate beim Pap Test können Qualitätskontrollen beitragen. Darüber hinaus befinden sich gegenwärtig in Deutschland neue Technologien der Dünnschichtpräparationen und der Computergestützten Auswertung von Zervixabstrichen in der Markteinführungsphase, die mit dem Ziel entwickelt wurden, die Rate der falsch-negativen Pap Tests zu senken.

Fragestellung

Lässt sich in diagnostischen Studien die Testgüte, also die Genauigkeit und Zuverlässigkeit der Dünnschichtpräparationen und der Computergestützten Auswertungen bestimmen und wie unterscheiden sich die genannten neuen Technologien in diesen Parametern vom konventionellen Pap Test? Gibt es darüber hinaus Erkenntnisse in der Literatur, die eine prognostische Aussage zulassen, inwieweit die neuen Verfahren Inzidenz und Mortalität des Zervixkarzinoms positiv beeinflussen?

Methodik

Zielpopulation sind alle Frauen ab dem 20. LJ. Als Ergebnisparameter zur Bestimmung der diagnostischen Genauigkeit und Zuverlässigkeit wurden Sensitivität, Spezifität, FPR, FNR, Likelihood Ratios, diagnostische Odds Ratios und, bei gleichzeitiger Angabe der Prävalenzdaten in der Studienpopulation, PPV und NPV gewählt. Studien wurden auch dann eingeschlossen, wenn sie über andere als die genannten Ergebnisparameter berichteten und eine nachträgliche Ermittlung aus den mitgeteilten Rohdaten möglich war, sofern ein adäquater Referenzstandard eingesetzt wurde. Für prognostische Studien wurden die Outcomeparameter Senkung der Inzidenz und Mortalität an Zervixkarzinom festgelegt. Im Februar 2001 wurde eine systematische Literaturrecherche, bestehend aus einer systematischen Suche in den elektronischen Datenbanken Medline, EMBASE, CancerLit und Science Citation Index, erweitert um eine Handsuche in den Datenbanken internationaler HTA-Institutionen, auf den Homepages von Leitlinienherstellern und Fachgesellschaften und eine Handsuche in den Publikationslisten bereits ermittelter Dokumente, durchgeführt, ein Update der Recherche erfolgte im November 2001. Hersteller von Geräten der hier evaluierten Technologien wurden in standardisierter Weise befragt. Für einen Einschluss in den vorliegenden Bericht mussten die identifizierten Dokumente definierte Kriterien erfüllen. Die Bewertung der methodischen Qualität von HTA-Berichten und Übersichtsarbeiten wurden anhand von Checklisten vorgenommen, primäre Studien wurden bezüglich ihrer methodischen Qualität in eine von vier Qualitätsstufen klassifiziert, gestützt auf ein modifiziertes, von Flynn & Adams 1996 vorgeschlagenes Schema. Die Synthese der Daten erfolgte narrativ, da die identifizierten Publikationen aufgrund einer erheblichen Heterogenität eine quantitative Synthese in Form einer Metaanalyse nicht zuließen.

Ergebnisse

Insgesamt wurden vier HTA-Berichte und eine Metaanalyse in die Bewertung der medizinischen Effektivität eingeschlossen, davon behandelten ein HTA-Bericht und

die Metaanalyse nur die Dünnschichtpräparationen. Weiterhin konnten in die Evaluation sechs primäre Studien zur diagnostischen Genauigkeit und Zuverlässigkeit aufgenommen werden. Darunter befanden sich keine Studien aus Deutschland. Es konnten keine prognostischen Studien identifiziert werden. Übereinstimmend wurde in allen HTA-Berichten eine unzureichende Evidenzlage für Testgütekriterien aller neuen Screeningtechnologien, d.h. sowohl für Dünnschichtpräparationen, als auch für Computergestützte Auswertungen von Pap Tests, sowie die überwiegend mangelhafte Qualität bei den durchgeführten klinischen Studien konstatiert. Für Dünnschichtpräparationen liegen Hinweise vor, dass diese Verarbeitungstechnik einen höheren Anteil adäquater Abstrichproben liefern kann. Die berücksichtigten Primärpublikationen wiesen überwiegend inadäquate Referenzstandards und/oder schwerwiegende methodische Mängel auf. Diese Qualitätsmängel begründeten die unzureichende interne und externe Validität der Ergebnisse. Die hier bewerteten Primärstudien über Dünnschichtpräparationen haben die Evidenzlage, wie sie in den HTA-Berichten beschrieben wurde, nicht verändert. Insbesondere ergaben sie keinen Erkenntniszuwachs hinsichtlich der Testgütekriterien. Auch bei der Bewertung des AutoPap™-Verfahrens, der Computergestützten Auswertung von Pap Tests, konnten aus den beiden identifizierten Studien keine validen Testgütekriterien gewonnen werden.

Schlussfolgerung

Aus den hier ausgewerteten Primärstudien und Kontextdokumenten konnte keine Evidenz für eine Überlegenheit computergestützter Auswertungen oder Dünnschichtpräparationen gegenüber konventionellen Pap Tests gewonnen werden. Die diagnostische Genauigkeit und Zuverlässigkeit der neuen Technologien sind noch nicht hinreichend genau bestimmt worden. Die Anwendung der Dünnschichtpräparationen ergibt Präparate, die sich von den mit konventionellen Pap Tests gewonnenen Abstrichen unterscheiden. Die Dünnschichtpräparationen können gleichzeitig einen Informationsverlust und eine bessere Auswertbarkeit der Präparate bedeuten. Die Hinweise für eine erhöhte Zahl adäquater Abstrichproben, die mit dieser Technologie gewonnen werden können, waren jedoch nicht konsistent, wenn gleichzeitig die Sorgfalt bei der Entnahmetechnik konventioneller Pap Abstriche mit adäquaten Abstrichinstrumenten und Beachtung wichtiger Grundregeln einbezogen wurde. Somit besteht auch auf der Stufe technischer Machbarkeit (Flynn; 1996) noch keine klare Evidenzlage. Das heißt, dass auch auf dieser Stufe die Dünnschichtpräparationstechnik der konventionellen Technologie nicht als überlegen eingestuft werden kann. Für die Computergestützte Auswertung von Pap Abstrichen mittels AutoPap™ liegen zwar Ergebnisse vor, die eine Überlegenheit als Rescreener gegenüber 10% manuellem Rescreening konventioneller Pap Abstriche zur Senkung der Falsch-Negativen Rate zeigen, eine Überlegenheit gegenüber anderen Verfahren, wie z.B.

dem manuellen Rapid Rescreening ist jedoch nicht belegt. Für den Einsatz von AutoPap™ als Primärscreener ergab sich ebenfalls keine neue Evidenz, aus der sich eine Überlegenheit der Computergestützten Auswertung von Pap Abstrichen hinsichtlich ihrer diagnostischen Genauigkeit und Zuverlässigkeit hätte ableiten lassen. Es besteht somit weiterer Forschungsbedarf, um präzise und valide Ergebnisse zur Testgenauigkeit und Zuverlässigkeit beider Technologien zu ermitteln, auch dahingehend, ob Dünnschichtpräparationstechniken und Computergestützte Auswertung in kombiniertem Einsatz eine Verbesserung der Testgütekriterien des konventionellen Pap Tests nach sich ziehen könnten. Da Verbesserungen der Screeningstrategie und der Screeningteilnahmeraten größere Auswirkungen auf die Steigerungen der Effektivität des Screeningprogramms unter Alltagsbedingungen hervorrufen könnten, als Verbesserungen der Sensitivität und Spezifität des einzelnen Tests, sollte unseres Erachtens ein prioritärer Forschungsschwerpunkt in der Untersuchung der Effektivität von Interventionen für die Steigerung der Teilnahmeraten beim Screening bestehen.

B.2 Ökonomische Evaluation

U. Siebert, G. Sroczynski, K. Voigt, P. Aidelsburger, J. Engel, B. Gibis, P. Hillemanns, D. Hölzel, J. Wasem

Einleitung

Das Zervixkarzinom ist weltweit die zweithäufigste tumorbedingte Todesursache der Frau. Im Jahr 1995 verstarben in Deutschland 2.207 Frauen an einem Zervixkarzinom, dies sind 2% aller Krebstodesfälle. Die standardisierte Mortalitätsrate lag bei 4,2 pro 100.000 Frauen und Jahr. Trotz bundesweiter Einführung von Krebsfrüherkennungsuntersuchungen im Rahmen des Krebsfrüherkennungsprogramms 1971 in Deutschland bleibt die jährliche Neuerkrankungsrate für das Zervixkarzinom mit 7000 Fällen pro Jahr hoch und liegt im europäischen Vergleich im oberen Drittel.

Die Zervixkarzinomfrüherkennung erfolgt durch eine gynäkologische Untersuchung sowie die Entnahme, Verarbeitung und zytologische Beurteilung eines Zervixzell-Abstrichs. Der zytologische Abstrich wird modifiziert nach dem Papanicolaou-Verfahren durchgeführt. Zunehmend wird im Fachkreis die Qualität und Effizienz dieses Screeningverfahrens kritisiert. Ferner werden neue Präparationstechniken und computergestützte Auswertungstechnologien als Screeningverfahren diskutiert, die zum einen die Qualität des Zellabstriches verbessern können, zum anderen falsch-negative Befundraten verringern sollen.

Nach den bestehenden gesetzlichen Grundlagen (§2 Abs. 1 SGB V) haben Qualität und Wirksamkeit der erbrachten Leistungen im Gesundheitswesen dem allgemein anerkannten Stand der medizinischen Erkenntnisse zu entsprechen und den medizinischen Fortschritt zu berücksichtigen. Das Wirtschaftlichkeitsgebot (§12 Abs. 1 SGB V) gebietet, dass die Leistungen „ausreichend, zweckmäßig und wirtschaftlich“ sein müssen und „das Maß des Notwendigen nicht überschreiten“ dürfen. Neue Untersuchungs- und Behandlungsmethoden in der ambulanten vertrags(zahn)ärztlichen Versorgung sind nur dann abrechnungsfähig, wenn der Bundesausschuss der Ärzte und Krankenkassen deren diagnostischen oder therapeutischen Nutzen sowie deren medizinische Notwendigkeit und Wirtschaftlichkeit - auch im Vergleich zu bereits zu Lasten der Krankenkassen erbrachten Methoden - anerkannt hat (§135 Abs.1 SGB V).

Ziel dieser ökonomischen Evaluation war die Beurteilung der ökonomischen Konsequenzen des Einsatzes der Dünnschichtpräparation und computergestützter Auswertungstechnologien beim Zervixkarzinom-Screening im Rahmen des gesetzlichen Krebsfrüherkennungsprogramms anhand der derzeit verfügbaren wissenschaftlichen Evidenz. Ferner sollte die Übertragbarkeit der Daten auf das deutsche Gesundheitssystem beurteilt und der zukünftige Forschungsbedarf identifiziert werden. Die Ergebnisse dieser Evaluation sollen zur gesundheitspolitischen Diskussion darüber, ob eine Aufnahme der untersuchten Screening-Technologien in den Leistungskatalog der gesetzlichen Krankenversicherung sinnvoll erscheint, einen klärenden Beitrag liefern.

Methoden

Es wurde eine systematische Literaturrecherche in den relevanten Zeitschriften- und HTA-Datenbanken durchgeführt: Medline, PreMedline, Pubmed, Embase, HealthStar, CINAHL, Cancerlit, DARE, NEED, EconLit, NHS HTA-Databases, Cochrane Library, Best Evidence und Web of Science (Science Citation Index, Social Science Citation Index and Arts & Humanities Citation Index). Es wurde eine deutsche und eine internationale Literatursuche mit einer Freitextsuche durch Schlagwörter und deren Verknüpfungen mit AND/OR-Operatoren durchgeführt. Die internationale Recherche beschränkte sich auf die Suche aktueller Publikationen, die in internationalen HTA-Berichten angegeben wurden. Internetseiten wurden nach relevanter (Hintergrund-) Information und epidemiologischen Daten, sowie Richtlinien und Empfehlungen durchsucht. Zusätzlich wurden die Referenzlisten der identifizierten Publikationen auf relevante Beiträge gesichtet. Ferner wurden nationale und internationale Experten im Bereich Epidemiologie, Gynäkologie, Zytologie, Gesundheitsökonomie, medizinische Entscheidungsanalyse und Ethik sowie Hersteller und Vertreiber der untersuchten Technologien in Deutschland kontaktiert und zum Thema befragt.

Es erfolgte eine systematische Literaturselektion durch zwei der Autoren des HTA-Teams anhand a priori festgelegter Ein- und Ausschlusskriterien. Bei Diskrepanzen wurde ein dritter Gutachter hinzugezogen und ein Konsensusentscheid herbeigeführt. Es wurden ausschließlich Volltext-Publikationen in deutscher oder englischer Sprache eingeschlossen. Berücksichtigte ökonomische Studien wurden anhand der standardisierten Checkliste der German Scientific Working Group Technology Assessment for Health Care zur Beurteilung der methodischen Qualität gesundheitsökonomischer Studien bewertet. Die Bewertung erfolgte zunächst durch zwei Projektmitarbeiter jeweils unabhängig voneinander. Im daran anschließenden Vergleich der Ergebnisse wurden Diskrepanzen im gemeinsamen Gespräch ausge-

räumt. Konnte kein Konsens erzielt werden, wurde wiederum ein dritter Bewerter hinzugezogen.

Die Daten der ökonomischen Dokumente wurden systematisch extrahiert und in einer Datenbank zusammengestellt. Es wurde dafür die vorgegebene Datenbankstruktur des gesundheitsökonomischen Panels der German Scientific Working Group Technology Assessment for Health Care verwendet. Die Ergebnisse wurden in einer qualitativen Synthese entsprechend den gestellten Forschungsfragen als tabellarische Synopsis zusammengestellt und die Übertragbarkeit auf das deutsche Gesundheitssystem beurteilt.

Die ökonomischen Experten der HTA-Arbeitsgruppe erhoben ökonomische Primärdaten wie Mengen- und Preisgerüste und führten diese zu Kostendaten direkter medizinischer Kosten für das Zervixkarzinom-Screening und die Zervixkarzinomtherapie in Deutschland zusammen. Basisjahr bei der Erhebung der Preise war das Jahr 2000. Alle Kosten wurden mit dem entsprechenden festen Umrechnungsfaktor von Deutscher Mark (DM) in Euro (€) umgerechnet (1 € entspricht 1,95583 DM). Zugrundeliegende Mengen basierten auf Angaben des Tumormanuals „Empfehlungen zur Diagnostik, Therapie und Nachsorge – Zervixkarzinom“ und wurden durch Expertenmeinung ergänzt. Die Bewertung der Mengen erfolgte durch modifizierte administrative Preise als Substitute für Marktpreise. Bei der ambulanten ärztlichen Versorgung wurden entsprechend der Empfehlungen der Arbeitsgemeinschaft „Reha-Ökonomie“ die ärztlichen Leistungen mit einem Durchschnittspreis basierend auf der Gebührenordnung für Ärzte und dem Einheitlichen Bewertungsmaßstab bewertet. Der Durchschnittspreis ergab sich unter Berücksichtigung des GKV/PKV-Versichertenverhältnisses. Dabei wurden den unterschiedlichen Preisen in alten und neuen Ländern und dem Verhältnis von in GKV und PKV versicherten Frauen Rechnung getragen. Die Daten zur Berechnung dieser Gewichtung wurden dem Mikrozensus des Jahres 1999 entnommen. Für computergestützte Auswertungsverfahren und Dünnschichtpräparationsverfahren wurde eine detaillierte Kostenaufstellung (microcosting) basierend auf Kostendaten, die von technologievertreibenden Institutionen in Deutschland zur Verfügung gestellt wurden, durchgeführt. Bei allen Analysen wurde die gesamtgesellschaftliche Perspektive eingenommen.

Ergebnisse

Von insgesamt 30 identifizierten relevanten ökonomischen Studien erfüllten 11 Studien die a priori vorgegebenen Einschlusskriterien für eine systematische Bewertung anhand der Checkliste zur methodischen Qualität gesundheitsökonomischer Studien. Es wurde keine deutsche Studie zur Kosten-Effektivität der untersuchten Screeningverfahren identifiziert.

In der ökonomischen Primärdatenerhebung für Deutschland wurden durchschnittliche Kosten für eine Zervixkarzinom-Früherkennungsuntersuchung bestehend aus der gynäkologischen Früherkennungsuntersuchung mit zytologischer Abstrichentnahme und der zytologischen Auswertung von 17,28 € ermittelt. Die Kosten für eine Therapie bei Auftreten eines Zervixkarzinoms waren stadienabhängig und variierten mit den aktuellen Therapierichtlinien und Empfehlungen. Die errechneten mittleren Gesamtkosten lagen zwischen 2.000 € und 9.051 €. Ferner fallen jährliche Medikationskosten für lokale und systemische Östrogene, Analgetika sowie regulierende Medikamente nach Therapie zwischen 216 € bis 749 € sowie jährliche Kosten der Nachsorgebehandlungen zwischen 2.007 € in den ersten beiden Jahren und 49 € ab dem sechsten post-operativen Jahr an.

Auf Grundlage der von den kontaktierten Herstellern bzw. Vertreibern zur Verfügung gestellten Kostendaten wurde eine Berechnung der Kosten pro Präparat für die in Deutschland vertretenen Technologien ThinPrep™, AutoCytePrep und AutoPap GS Screening System durchgeführt. Die Kostenberechnungen der verschiedenen Technologien basierten auf unterschiedlichen Rechenansätzen und unterschiedlich detaillierter Datenlage. Ein direkter Vergleich der verschiedenen Technologien ist aus diesem Grund nur schwer möglich. Die Kosten lagen zwischen 4€ und 19€ pro Präparat. In internationalen Kosten-Effektivitäts-Studien wurden im Vergleich zum konventionellen Papanicolaou-Test inkrementelle Kosten im Bereich von 3–30 US\$ pro Test für neue Screening-Verfahren angegeben. Dabei umfassen die angegebenen Kosten die Material- und Durchführungskosten pro Zellabstrich.

In internationalen entscheidungsanalytischen Studien betrug das diskontierte inkrementelle Kosten-Effektivitäts-Verhältnis (IKEV) bei einem 3-Jahres-Screeningintervall für die Dünnschichtpräparation 10.000-37.000 US\$ pro gewonnenem Lebensjahr (LJ), für automatisierte Primärscreening-Verfahren 22.000-29.000 US\$/LJ und für automatisierte Rescreening-Verfahren 16.000-48.000 US\$/LJ, jeweils im Vergleich zum konventionellen Papanicolaou-Verfahren mit manueller Auswertung. Kürzere Screeningintervalle führten jeweils zu einer höheren diskontierten IKEV. Sensitivitätsanalysen zeigten, dass die Modellergebnisse einer Reihe von Unsicherheiten in wichtigen Modellparametern insbesondere bei den Testgütekriterien und Testkosten unterlagen. Alle eingeschlossenen ökonomischen Studien wiesen Limitierungen auf. Keine der Studien hatte die Gesundheitseffekte mit Präferenzen bewertet und bei der Kostenerhebung indirekte Kosten berücksichtigt. Kosten für die neuen Technologien berücksichtigten in der Regel keine Personalschulungskosten, Lagerungskosten und Transportkosten. Die Werte für Sensitivität und Spezifität der Testverfahren unterlagen einer großer Variabilität und mussten oft wegen fehlender Evidenz von Experten geschätzt werden. Häufig wurde die Annahme gemacht, dass sich die marginale Erhöhung der Sensitivität der neuen Technologie bei allen Schwellenwerten gleich

ausprägt. Aufgrund fehlender Evidenz wurde für die Spezifität neuer Technologien häufig die Spezifität des konventionellen Pap-Tests verwendet.

Diskussion

Es wurden keine randomisierten kontrollierten Studien oder Feldstudien identifiziert, die den ökonomischen Aspekt eines Einsatzes neuer Screeningtechnologien untersucht haben. Alle berücksichtigten gesundheitsökonomischen Studien basierten auf entscheidungsanalytischen Markov-Modellen, die den Krankheitsverlauf simulierten.

Entscheidungsanalytische Markov-Modelle bieten einen systematischen, expliziten und quantitativen Ansatz, um wissenschaftliche Evidenz zu diagnostischer Testgüte und Langzeitprognose zu verknüpfen und diese wiederum mit ökonomischen Parametern zusammenzuführen. Ferner können neben den Zielgrößen Lebenserwartung und Lebenszeitkosten weitere für die Entscheidungsfindung und Bedarfsplanung wichtige klinische Outcomes evaluiert werden, wie beispielsweise zusätzlich entdeckte Krebsfälle, vermiedene Krebsfälle, vermiedene Todesfälle etc. Die erforderlichen Annahmen eines entscheidungsanalytischen Modells können im Rahmen von Sensitivitätsanalysen modifiziert und ihr Einfluss auf das Analyseergebnis untersucht werden. Dieses ermöglicht die Beurteilung der Robustheit der Modellergebnisse und außerdem die Identifikation unsicherer Parameter, für die weiterer Forschungsbedarf besteht.

Alle internationalen Kosten-Effektivitäts-Studien zeigten, dass ein Einsatz der neuen Screeningtechnologien im Primär- bzw. Rescreening kosteneffektiv sein kann, wenn das Screening in Zeitabständen von drei oder mehr Jahren durchgeführt wird. Kürzere Screeningintervalle erhöhen jedoch das Kosten-Effektivitäts-Verhältnis. Die Studienergebnisse der eingeschlossenen Studien unterlagen einer breiten Streuung.

Je kürzer das Screeningintervall, um so teurer war der Einsatz eines Screeningverfahrens. Die Ursachen hierfür sind vergleichsweise häufiger durchgeführte Screeningtests und somit ein höherer Anteil erkannter Dysplasien und daraus resultierende diagnostische Abklärungen und Behandlungen. Bei den erkannten Krebsvorstufen handelt es sich jedoch häufig um geringgradige Veränderungen, die sich ohne Behandlung zum großen Teil spontan zurückbilden bzw. bei Persistenz auch bei einem beispielsweise 3-jährigen Screeningintervall noch rechtzeitig erkannt und behandelt werden könnten, da beim Zervixkarzinom eine relativ langsame Progression stattfindet.

Bei der derzeitigen jährlichen Screeningpraxis in Deutschland ist ein Einsatz der neuen Technologien in der Zervixkarzinomfrüherkennung nicht als kosteneffektiv zu bewerten. Aus gesundheitsökonomischer Sicht wäre allerdings der Einsatz neuer Technologien bei einer gleichzeitigen Verlängerung des Screeningintervalls auf 3 Jahre zu diskutieren. Letztlich hängt die Länge des Screeningintervalls mitunter von der gesellschaftlichen Zahlungsbereitschaft (willingness-to-pay) ab.

Die Sensitivitätsanalysen der untersuchten Studien zeigten, dass die Modellergebnisse einer Reihe von Unsicherheiten in wichtigen Modellparametern unterliegen. Die Ergebnisse bezüglich der Kosten-Effektivität der neuen Screeningtechnologien waren in allen Studien deutlich sensitiv für Änderungen in den Testgütekriterien Sensitivität und Spezifität. Gerade für diese Modellparameter war jedoch die Evidenzgrundlage mangelhaft. In Abwesenheit zuverlässiger und valider Daten wurde in den meisten untersuchten Modellanalysen die Annahme gemacht, dass die Sensitivität der neuen Testverfahren im Vergleich zur herkömmlichen Methode höher liegt und diese Erhöhung für alle Dysplasiegrade (CIN-Stadien) gleich hoch ist. Die Annahme einer konstanten Verbesserung der Sensitivität unabhängig vom Abnormalitätsgrad spiegelt jedoch nicht unbedingt die Realität wider. Einzelne Studien zur medizinischen Effektivität der neuen Screeningverfahren zeigten z.B. keine oder nur eine geringere Erhöhung der Sensitivität für hochgradige Dysplasien im Vergleich zur konventionellen Screeningmethode (Broadstock 2000). Überschätzte inkrementelle Sensitivitätserhöhungen für hochgradige Dysplasien für die neuen Screeningtechnologien können zu fehlerhaften Modellrechnungen für die gewonnene Restlebenserwartung durch ein Screening mit der neuen Technologie versus einem Screening mit der konventionellen Methode führen. Ferner wurde in vielen der untersuchten Studien aufgrund mangelhafter Evidenz angenommen, dass sich die Spezifität der neuen Verfahren nicht von der Spezifität des konventionellen Screeningverfahrens unterscheidet. Eine Erhöhung der Sensitivität kann jedoch aufgrund von Cutpoint-Effekten auch mit einer Reduktion der Spezifität einhergehen. Eine tatsächliche Erhöhung der falsch-positiven Befundrate bei neueren Technologien hätte einen erheblichen Einfluss auf Kosten und ggf. auch auf die Lebensqualität oder gar Ereignisse wie Konisationen bei Patientinnen mit wiederholt falsch-positiven Testergebnissen. Dies wiederum kann zu deutlichen Erhöhungen der inkrementellen Kosten-Effektivitäts-Verhältnisse führen.

Die Ergebnisse der Modellanalysen unterlagen einer breiten Streuung gegenüber Änderungen in den Kosten der neuen Technologien. Testkosten basierten häufig ausschließlich auf Verbrauchsmaterial- und Kapitalkosten. Kosten für den Geräteservice, Raumnutzung, Mitarbeitertraining, Durchführung der Tests, Transport und Lagerung blieben meist unberücksichtigt, so dass von einer Unterschätzung der tatsächlichen Testkosten auszugehen ist. Kosten für administrative Tätigkeiten kön-

nen bei den neuen computergestützten Verfahren durchaus höher sein als bei der konventionellen Methode. Kosten für die Testdurchführung sind abhängig von der benötigten Zeit pro Testung und sind aus diesem Grund oftmals schwierig objektiv zu bestimmen. Für große Labore mit großer Kapazität können sich jedoch die Kosten pro Einheit reduzieren.

Qualitätskorrigierte Lebensjahre (QALY) sind ein bevorzugtes Effektivitätsmaß einer gesundheitsökonomischen Evaluation. Da die erforderlichen Nutzwerte für die verschiedenen Krankheitszustände in der Literatur fehlen, wurde die Lebensqualität in keiner der eingeschlossenen Studien berücksichtigt. Es ist jedoch davon auszugehen, dass ein positives Testergebnis negative psychologisch-traumatische Auswirkungen auf einen Teil der gescreenten Frauen hat, die beispielsweise durch die Angst vor bevorstehenden Untersuchungen, Diagnosen und Behandlungen verursacht werden. Eine Nichtberücksichtigung der Lebensqualitätsaspekte führt demnach zur Unterschätzung der tatsächlichen inkrementellen Kosten-Effektivitäts-Verhältnisse.

Ein zukünftiger Forschungsbedarf liegt somit in der Erhebung valider Lebensqualitätsdaten in Abhängigkeit vom Krankheitsstadium der Zervixkarzinom-erkrankung und in Abhängigkeit vom Screeningtestergebnis. Ferner ist dringend erforderlich, dass valide Erhebungen der Testgüte neuer Verfahren bei der Durchschnittsbevölkerung (ohne Risikoselektion) unter Routineverhältnissen und im Vergleich zu einem Goldstandard durchgeführt werden. Aus ethischen Gründen sollte hierbei das Goldstandard-Follow-Up nur bei einem Teil der testnegativen Frauen durchgeführt werden. Diese Maßnahme ermöglicht die statistische Korrektur eines Verification Bias, der entsteht, wenn (falsch-)testnegative Patientinnen keiner konfirmatorischen Goldstandard-Untersuchung unterzogen werden. Ein weiterer wichtiger Forschungsbedarf liegt in der Untersuchung von Möglichkeiten zur Verbesserung der regelmäßigen Teilnahme am Screening, insbesondere in unterversorgten Subpopulationen mit niedrigen Screeningteilnahmeraten und im Hinblick auf eine mögliche Verlängerung des Screeningintervalls.

Die evaluierten gesundheitsökonomischen Studien stammen aus Australien, Großbritannien, Kanada und den USA. Ob und in wieweit eine Übertragbarkeit dieser Studienergebnisse auf das Gesundheitssystem in Deutschland möglich ist, ist fraglich. Demographie, Epidemiologie der Erkrankung, klinische Praxis und Richtlinien sowie Struktur des Gesundheits- und Vergütungssystems, Preise und Verbraucherpräferenzen in Deutschland differieren von den anderen Ländern. Aufgrund der erheblichen Heterogenität der Ergebnisse der einzelnen Studien erscheint eine direkte Übertragbarkeit auf den deutschen Kontext nicht realistisch und notwendige Anpassungsverfahren sind in vielen Bereichen aufgrund mangelnder Datenlage nicht

möglich. Eine Alternative, die eine valide Bewertung der Langzeiteffektivität und Kosteneffektivität der neuen Screeningtechnologien im deutschen Gesundheitssystem ermöglichen würde, liegt in der Entwicklung und Auswertung eines für Deutschland spezifischen entscheidungsanalytischen Modells mit Lebenszeithorizont.

Schlussfolgerungen und Empfehlungen

Basierend auf internationalen Kosten-Effektivitäts-Studien kann der Einsatz neuer Technologien beim Zervixkarzinom-Screening zu akzeptablen Kosten pro gewonnenem Lebensjahr führen, wenn das Screening in einem Zeitintervall von drei oder mehr Jahren erfolgt. Für den derzeitigen deutschen Standard des jährlichen Screenings ist die Einführung der neuen Screeningverfahren als nicht kosteneffektiv zu bewerten.

Die bestehenden Unsicherheiten im Bereich der Effektivität und Kosten der neuen Verfahren erschweren jedoch eine valide Aussage zur Bestimmung einer optimalen Screening-Strategie und eines optimalen Screeningintervalls.

Zukünftiger Forschungsbedarf besteht in validen Erhebungen der Testgüte neuer Verfahren in der Allgemeinbevölkerung unter Verwendung eines Goldstandards. Ferner sind Möglichkeiten der Verbesserung der Teilnahme der Frauen am regelmäßigen Screening, insbesondere in unterversorgten Subpopulationen mit mangelhafter Screeningteilnahme und im Hinblick auf eine mögliche Verlängerung des Screeningintervalls, zu untersuchen. Weiterer Forschungsbedarf liegt in der Erhebung valider Lebensqualitätsdaten von Frauen in Abhängigkeit von verschiedenen Krankheitsstadien (HPV-Infektion, Dysplasiestufen, Karzinomstufen) und Screeningtestergebnissen.

Die Entwicklung eines entscheidungsanalytischen Modells für den deutschen Versorgungskontext ist notwendig, um die Langzeit-Effektivität und Kosten-Effektivität neuer Untersuchungsmethoden von Zervixabstrichen im Kontext des deutschen Gesundheitswesens evaluieren zu können.

B.3 Systematische Entscheidungsanalyse

U. Siebert, G. Sroczynski, B. Gibis, P. Aidelsburger, J. Engel, P. Hillemanns, K. Voigt, C. Muth, M. Velasco-Garrido, J. Wasem, D. Hölzel, S.J. Goldie

Einleitung

Obwohl seit 1971 in Deutschland Krebsfrüherkennungsuntersuchungen in den Leistungskatalog der gesetzlichen Krankenversicherung (GKV) einbezogen und als Krebsfrüherkennungsprogramm bundesweit eingeführt wurden (§ 25, Abs. 2, SGB V, 2000b), liegt die Zervixkarzinominzidenz mit 7000 Neuerkrankungen pro Jahr in Deutschland im europäischen Vergleich im oberen Drittel.

Der Effektivität der Screeningverfahren und der eingeleiteten therapeutischen Maßnahmen sowie dem Screeningintervall und der Beteiligungsrate am Screening kommen deutliche Effekte auf die Gesamt-Effektivität und –Effizienz der Zervixkarzinomfrüherkennung zu. Derzeit wird in Deutschland – wie in den meisten westlichen Ländern – das Verfahren nach Papanicolaou durchgeführt. Eine kürzlich publizierte Meta-analyse berichtete jedoch eine relativ geringe Sensitivität dieses Screeningverfahrens. Neuere Verfahren, wie die Dünnschichtpräparation und die (halb-)automatisierten Auswertungsverfahren zytologischer Abstriche erzielten in einzelnen klinischen Studien eine höhere Sensitivität. Derzeit ist der Einsatz dieser modernen Technologien jedoch teuer und seine Effektivität umstritten.

Da entsprechende randomisierte Studien fehlen, stützten sich Empfehlungen zum Zervixkarzinom-Screening insbesondere in den USA bislang auf entscheidungsanalytische Modelle. In diesen Modellen wurden epidemiologische Daten über den natürlichen Krankheitsverlauf des Zervixkarzinoms, die Textcharakteristika der Zervixkarzinom-Früherkennungsverfahren und Kostendaten zusammengeführt, um den Einfluss von Interventionen quantitativ ermitteln zu können. Die Übertragbarkeit dieser Ergebnisse auf das deutsche Gesundheitssystem bleibt jedoch auf Grund von Unterschieden in der Epidemiologie, der diagnostischen und therapeutischen Praxis und den Kosten fraglich.

Aus diesem Grund wurde ein entscheidungsanalytisches Modell für das deutsche Gesundheitssystem entwickelt und analysiert. Das Ziel der systematischen Entscheidungsanalyse war die Ermittlung der medizinischen und ökonomischen Langzeitkon-

sequenzen des Einsatzes neuer Screeningtechnologien in der Zervixkrebsfrüherkennung unter Einbeziehung der derzeitigen wissenschaftlichen Evidenz und der für Deutschland spezifischen klinischen und gesetzlich festgelegten Praxis. Ferner sollte die Bedeutung von Unsicherheiten in den relevanten Faktoren festgestellt werden. Es wurde eine klinische Entscheidungsanalyse und eine Kosten-Effektivitäts-Analyse zum Einsatz verschiedener Screeningtechnologien in der Zervixkarzinomfrüherkennung für den deutschen Versorgungskontext durchgeführt. Die Ergebnisse sollen die Entscheidungsträger in Deutschland bei der rationalen Allokation von Ressourcen unterstützen.

Methoden

Design

Es wurde ein entscheidungsanalytisches Markov-Modell entwickelt, das den natürlichen Krankheitsverlauf und den Krankheitsverlauf unter den verschiedenen Screeningstrategien mit den jeweils anfallenden Kosten für Screening, Diagnostik und Behandlung einer Kohorte von Frauen im Alter von 15 Jahren über deren Lebenszeit abbildet. Dabei wurden deutsche epidemiologische Daten, Angaben zur aktuellen deutschen Praxis in Diagnose und Behandlung des Zervixkarzinoms und seiner Vorstufen sowie deutsche Kostendaten verwendet. Frauen konnten im Modell ab einem Alter von 20 Jahren an einem jährlichen Screening teilnehmen. In der Basisfallanalyse wurden Frauen mit regelmäßiger Teilnahme am Screening untersucht.

Für die folgenden Strategien wurden medizinische Effektivität, Lebenszeitkosten und Kosten-Effektivität ermittelt: (1) Kein Screening, (2) Papanicolaou-Verfahren plus manuelle zytologische Auswertung, (3) Dünnschichtpräparation plus manuelle zytologische Auswertung, (4) Papanicolaou-Verfahren plus automatisierte zytologische Auswertung, (5) Dünnschichtpräparation plus automatisierte zytologische Auswertung.

Zielgrößen der Analysen waren entdeckte und vermiedene Karzinomfälle, zervixkarzinomspezifische Mortalität, die Lebenserwartung, Lebenszeitkosten und die Kosten-Effektivität verschiedener Screeningstrategien für Deutschland.

Beschreibung des Modells

Im entscheidungsanalytischen Modell konnten die Frauen basierend auf den zytologisch-histologischen Stadien der Zervixerkrankung verschiedene Gesundheits-

zustände annehmen: Keine Zervixzellläsion (Gesund), Zustand nach benigner Hysterektomie (Hysterektomie), milde zervikale intraepitheliale Neoplasien (CIN 1), moderate zervikale intraepitheliale Neoplasien (CIN 2), schwere zervikale intraepitheliale Neoplasien / Carcinoma in situ (CIN 3), invasives Zervixkarzinom FIGO-Stadium I bis IV (FIGO I-IV), Zustand Überlebend nach 5 Jahren post Diagnose eines Zervixkarzinoms (Survivor), Tod durch Zervixkrebs (Tod), Tod aus anderer Ursache (Tod). Frauen konnten aufgrund eines invasiven Zervixkarzinoms basierend auf krebstadienspezifischen Überlebenswahrscheinlichkeiten oder aus anderer Ursache in Abhängigkeit des Alters sterben, sie konnten in ihrem Gesundheitszustand verweilen oder zu einem anderen Gesundheitszustand progredieren oder regredieren. Eine Regression vom Gesundheitszustand des invasiven Karzinoms wurde im Modell ausgeschlossen. Eine Frau konnte in jedem Zustand mit Ausnahme der Stadien mit invasivem Karzinom eine benigne Hysterektomie erfahren und trat dann in den Zustand "Hysterektomie" ein, in dem kein spezifisches Screening für Zervixkarzinom mehr stattfand. Sowohl zervikale intraepitheliale Neoplasien als auch invasive Karzinome konnten durch ein auffälliges Screeningtestergebnis entdeckt und gemäß der deutschen klinischen Praxis kontrolliert oder behandelt werden. Unabhängig vom Screening konnte ein Zervixkarzinom aufgrund von klinischen Symptomen entdeckt werden.

Es wurden die folgenden Modellannahmen gemacht:

- Angaben in der Bethesda Klassifikation wurden wie folgt zugeordnet: Low-Grade Squamous Intraepithelial Lesion (LG-SIL oder LSIL) in CIN 1, High-Grade Squamous Intraepithelial Lesion (HG-SIL oder HSIL) in CIN 2, CIN 3 und Carzinoma in situ.
- Die Klassifikationsgruppe Atypical Squamous Cells of Undetermined Significance (ASCUS) ist nicht in die Münchner Nomenklatur zu übertragen und ist ein zytologischer und kein pathologischer Befund. ASCUS-Befunde wurden als negative Testbefunde gewertet. Alle im Modell verwendeten Werte für Sensitivität und Spezifität repräsentieren den zytologischen Grenzwert LSIL und höher.
- Eine Zervixkarzinomentwicklung kann nur ausgehend von CIN 3 im Anschluss einer Progressionsabfolge von CIN 1 über CIN 2 nach CIN 3 erfolgen. Eine schnelle Progression zum invasiven Karzinom ausgehend von CIN 1 oder CIN 2 innerhalb von weniger als 2-3 Jahren wurde nicht berücksichtigt.

- Behandelte Läsionen, bei denen kein invasives Zervixkarzinom vorliegt, gelten als geheilt. Im Modell erfolgt ein Übergang der entsprechenden Patientinnen in den Gesundheitszustand "Keine Läsion". Damit besitzen diese bereits behandelten Frauen dasselbe CIN- und Zervixkarzinomrisiko wie bislang nicht behandelte Frauen.
- Unabhängig vom zytologischen Testbefund wurde angenommen, dass ein Zervixkarzinom bei einer Frau, die am Screening teilnimmt, allein durch die gynäkologische Untersuchung via Tastbefund oder Inspektion entdeckt werden kann. Diese Entdeckungsrate wurde basierend auf Expertenschätzungen für die Stadien FIGO I, FIGO II, FIGO III und FIGO IV auf 70%, 75%, 80% und 85% festgelegt.
- An Zervixkarzinom erkrankte und behandelte Patientinnen haben fünf Jahre nach der Diagnose bzw. ihrem Therapiebeginn gegenüber gesunden Frauen keine erhöhte Mortalität mehr, sondern es gilt die allgemeine altersspezifische Hintergrundmortalität der deutschen weiblichen Allgemeinbevölkerung.
- Aufgrund der mangelhaften Datenlage zur Spezifität der Testverfahren wurden für die verschiedenen Testverfahren keine Unterschiede in der Spezifität berücksichtigt.
- Es wurde eine 100%ige Sensitivität für die Kombination von Kolposkopie mit Biopsie bei einer histologischen Abklärung einer intraepithelialen Läsion angenommen.

Natürlicher Krankheitsverlauf

Das Zervixkarzinom entwickelt sich langsam aus seinen Vorstadien, den intraepithelialen Neoplasien über einen Zeitraum von mehreren Jahren. In der Literatur werden Latenzzeiten vom Auftreten der Präkanzerose bis zum invasiven Zervixkarzinom von zehn Jahren und länger angegeben. Dabei ist im Stadium von CIN 1 und CIN 2 bei etwa der Hälfte der Fälle mit einer spontanen Rückbildung zu rechnen.

Daten zur Progression, Regression und Entdeckungsraten wurden aus der internationalen Literatur entnommen. Diese Daten wurden an deutsche Krebsregisterdaten für eine ungescreente Population aus den Jahren 1960 bis 1964 angepasst und das Modell somit für die Zervixkarzinom-Epidemiologie in Deutschland kalibriert. Die jährlichen Überlebenschancen bei einem Zervixkarzinom basierten auf

deutschen Tumorregisterdaten des Zeitraums 1988-1997. Basierend auf Daten des Statistischen Bundesamtes wurden altersspezifische Sterberaten der deutschen weiblichen Allgemeinbevölkerung herangezogen, um die Mortalität durch andere Ursachen zu modellieren.

Diagnostische Güte und therapeutische Wirksamkeit

Die Daten zur Sensitivität und Spezifität des Papanicolaou-Testverfahrens und den neuen Technologien wurden aus internationalen Studien entnommen. Grundlage hierfür bildete die im medizinischen und ökonomischen Teil dieses HTA-Reports eingeschlossenen internationalen HTA-Berichte und Primärstudien. Für den Pap-Test wurde eine Spezifität von 95% sowie eine Sensitivität von 47% für CIN1 und 72% für CIN2/3 aus einem publizierten systematischen Review verwendet. Für die Daten zur Testgüte der neuen Technologien wurde ein fallzahlgewichtetes Pooling durchgeführt und als Effektmaß die Richtig-Positiv Odds-Ratio (OR) berechnet. Diese wurde im entscheidungsanalytischen Modell multiplikativ auf die Sensitivität des Pap-Test angewandt, um die Sensitivität der neuen Strategie zu erhalten. Die berechneten Richtig-Positiv Odds-Ratios lagen zwischen 2,0 und 3,9.

Validierung

Das entscheidungsanalytische Modell wurde sowohl intern durch einen Vergleich der Modellprädiktionen mit den epidemiologischen Daten aus dem Gesamtkrebsregister Berlin als auch extern anhand von veröffentlichten deutschen Daten aus der Literatur und epidemiologischen Daten aus dem Krebsregister Saarland validiert.

Das Modell prädizierte eine mittlere Gesamtprävalenz CIN 1-3 und CIS von 7,2%. Der Altersgipfel der Dysplasien lag bei 27 Jahren. Das Lebenszeitrisiko einer Frau, an einem invasiven Zervixkarzinom zu erkranken, wurde vom Modell auf 3,0% prädiziert und liegt innerhalb der Spannweite anderer internationaler Publikationen von 2,5-3,3%. Daten des Krebsregisters Saarland für den Zeitraum 1968-1970 ergaben ein Risiko von 3,3%, bis zum 85. Jahr an Zervixkarzinom zu erkranken. Das vom Modell prädizierte Lebenszeitrisiko einer Frau für ein benigne Hysterektomie aus anderer Ursache als Zervixkrebs lag bei 30,5% und stimmte damit mit dem Anteil von 30% überein, der sich aus Daten des Frauengesundheitsberichts Bremen 2001 ergab.

Kosten

Es wurden direkte medizinische Kosten für Screeningtests, confirmatorische diagnostische Tests, ambulante und stationäre Versorgung, Arzneimitteltherapie, operative Therapien und palliative Maßnahmen berücksichtigt. Quellen für die Mengengerüste waren vorwiegend die Empfehlungen zur Diagnostik, Therapie und Nachsorge des Tumorzentrums München, die durch Expertenschätzungen ergänzt wurden. Wo erforderlich, wurde auf administrative Preise zurückgegriffen. Die Kosten des konventionellen Screeningverfahren wurden basierend auf den Abrechnungsziffern GOÄ Nr. 27 und EBM Nr. 157 für die gynäkologische Untersuchung und Abstrichentnahme gemäß der gesetzlichen Früherkennungsuntersuchung sowie den Ziffern GOÄ Nr. 4851 und EBM Nr. 4951 für die Färbung und zytologische Bewertung des Abstrichs berechnet. In der Berechnung der Kosten der neuen Screeningverfahren wurden die Kosten für die gynäkologische Untersuchung einschließlich Abstrichentnahme sowie die Kosten für die Färbung und zytologische Bewertung übernommen und aus der Mikrokostenberechnung die zusätzlichen Kosten für die Durchführung der neuen Verfahren, bestehend aus Materialkosten sowie Kapital-, und Instandhaltungskosten unter Berücksichtigung von Abschreibungen und einer jährlichen Verzinsung von 3%, hinzu addiert.

Es wurde eine gesamtgesellschaftliche Perspektive eingenommen. Das Basisjahr der Erhebung war das Jahr 2000. Alle Kosten wurden mit dem entsprechenden festen Umrechnungsfaktor von Deutscher Mark (DM) in Euro (€) umgerechnet (1 € entspricht 1,95583 DM). Kosten und Lebenserwartung wurden mit einer jährlichen Rate von 3% diskontiert. Da für Deutschland nach Kenntnis der Autoren kein Schwellenwert für die Kosten-Effektivität medizinischer Technologien im Sinne einer gesellschaftlichen Zahlungsbereitschaft publiziert ist, wurde in Anlehnung an US-Studien eine Technologie als kosteneffektiv bezeichnet, wenn das Kosten-Effektivitäts-Verhältnis unter 50.000 €/LJ lag.

Subgruppen- und Sensitivitätsanalysen

Für alle Modellparameter wurden Ein-Weg-Sensitivitätsanalysen unter Berücksichtigung der in der Literatur publizierten Spannweiten bzw. der 95%-Konfidenzintervalle durchgeführt. Einzelkosten wurden halbiert und verdoppelt. Die Diskontrate wurde zwischen 0% und 10% variiert. Die Werte für die Sensitivität der Screening-Testverfahren wurden basierend auf den zugehörigen 95%-Konfidenzintervallen der gepoolten Schätzer variiert. In einer Zwei-Weg-Sensitivitätsanalyse wurden simultan die Kosten des Screeningtests und die Sensitivität der neuen Screening-Verfahren variiert.

Es wurden neben der Basisfall-Analyse mit einem Screeningintervall von einem Jahr weitere Szenarien mit unterschiedlichen Screeningintervallen von 2, 3 und 5 Jahren analysiert. Die untere Altersgrenze zum Einschluss in das Screeningprogramm wurde zwischen 15 und 35 Jahren variiert. Die Teilnehmerate am Screening wurde zwischen 0% (keine Teilnahme der Frauen am Screening) bis 100% (vollständige regelmäßige Teilnahme aller Frauen am Screening) variiert. Es wurde dabei ausschließlich eine zufällige und von anderen Faktoren unabhängige Teilnehmerate pro Screeningzyklus verwendet.

Ergebnisse

Medizinische Effektivität

Die Modellrechnung ergab, dass in einer Kohorte von 100.000 Frauen ohne Screening 3.032 Frauen in ihrer Lebenszeit an einem Zervixkarzinom erkranken und dieses diagnostiziert wird. Das konventionelle jährliches Screening verhindert 2.994 Krebsfälle/100.000 Frauen, ein Umstieg von dem konventionellen auf eine der neuen Strategien würde zwischen 32 und 35 weitere Krebsfälle pro 100.000 Frauen verhindern. Gegenüber der Strategie kein Screening verhindert konventionelles Screening 1.000 zervixkarzinombedingte Todesfälle pro 100.000 Frauen. Mit den neuen Screeningverfahren würden 3-4 zusätzliche Todesfälle pro 100.000 Frauen verhindert werden.

Konventionelles Screening mit der Papanicolaou-Methode und manueller Auswertung erhöht die Lebenserwartung gegenüber "Kein Screening" um 94 Tage. Ein jährliches Screening mit den neuen Technologien würde im Vergleich zum konventionellen Screening die Lebenserwartung um etwa einen halben Tag erhöhen.

Kosten

Ohne Screening liegt der aus dem Modell berechnete Erwartungswert der undiskontierten absoluten Lebenszeitkosten bei 630 € pro Frau. Bei jährlichem Screening mit konventioneller Methode sind dies 1.220 €, mit automatisierter Auswertung 1.350 €, für die Dünnschichtpräparation plus manuelle Auswertung 1.970 € und für die Dünnschichtpräparation plus automatisierte Auswertung 2.070 €. Die Lebenszeitkosten pro Frau erhöhen sich damit bei einem Screening mit den neuen Verfahren im Vergleich zum konventionellen Verfahren je nach Verfahren zwischen 130 und 840 € bzw. zwischen 11% und 69%.

Kosteneffektivität

Im Vergleich zu keinem Screening hat ein jährliches Screening mit der konventionellen Papanicolaou-Methode ein diskontiertes inkrementelles Kosten-Effektivitäts-Verhältnis (IKEV) von 6.600 € pro gewonnenem Lebensjahr (€/LJ). Das IKEV liegt für die computergestützte Auswertung von Papanicolaou-Abstrichen bei 220.000 €/LJ, für die Dünnschicht-Präparation bei 953.000 €/LJ und für die Kombination von Dünnschicht-Präparation mit computergestützter Auswertung im Vergleich zum konventionellen Verfahren bei 1.083.000 €/LJ. Bei einem 2-Jahres-Intervall reduziert sich das IKEV des Screenings mit computergestützter Auswertung des Papanicolaou-Tests im Vergleich zum konventionellen Screening auf 26.000 €/LJ. Das IKEV des Screenings mit Dünnschicht-Präparation mit manueller bzw. computergestützter Auswertung reduziert sich auf 46.000 bzw. 53.000 €/LJ beim 3-Jahres-Intervall. Bei einer angenommenen Grenze der gesellschaftlichen Zahlungsbereitschaft von 50.000 €/LJ könnten die neuen Technologien bereits ab einem 2 bzw. 3-Jahres-Screening-Intervall als kosteneffektiv zu bewerten sein. Bei einem zweijährigen Screeningintervall sind die neuen Technologien kaum weniger effektiv als das jährliche konventionelle Screening, verursachen jedoch zum Teil deutlich weniger Kosten.

Subgruppen- und Sensitivitätsanalysen

In den Sensitivitätsanalysen blieben die qualitativen Aussagen der Basisfallanalyse weitgehend stabil. Einflussnehmende Variablen waren die Teilnehmerate am Screening, die Testgüte der neuen Technologien und die Testkosten sowie das Alter beim ersten Screening.

Mit höherer unterer Screening-Altersgrenze werden die IKEV zunächst günstiger, steigen jedoch bei einem Screeningbeginn in höherem Alter wieder an. Das konventionelle Screening ist für die gesamte Spannweite der unteren Altersgrenze kosteneffektiv. Bei Screeningbeginn im Alter von ca. ≥ 35 Jahren liegt die IKEV des Papanicolaou-Verfahrens mit automatisierter Auswertung unterhalb von 50.000 €/LJ und kann als kosteneffektiv bewertet werden.

Bei der Variation der zufälligen Teilnehmerate zeigte sich, dass das konventionelle Screeningverfahren im Vergleich zu "Kein Screening" bei jeder Teilnehmerate als kosteneffektiv zu bewerten ist. Für die computergestützte Auswertung des Papanicolaou-Tests fiel das IKEV bei einer Teilnehmerate von unter 65% unter den Schwellenwert von 50.000 €/LJ. Eine Einführung der Dünnschichtpräparation allein oder in Kombination mit einer computergestützten Auswertung würde diese Schwelle erst bei einer Teilnehmerate von weniger als circa 40% erreichen.

Werden simultan Testkosten und die Werte für die Sensitivität der neuen Screeningverfahren variiert, zeigt sich für das jährliche Screening, dass auch bei günstigen Werten für die Testkosten und die Sensitivität der neuen Screeningtechnologien keines dieser Verfahren ein akzeptables Kosten-Effektivitäts-Verhältnis erreicht. Selbst unter Annahme der für das IKEV günstigsten Werte der jeweiligen Testkosten und Sensitivität innerhalb der variierten Spannweite liegen die IKEV für alle neuen Screeningtechnologien über 170.000 €/LJ. Im Worst-Case-Szenario dieser Zwei-Weg-Sensitivitätsanalyse wurde die computergestützte Auswertung des Papanicolaou-Tests sogar dominiert, weil sie unter Berücksichtigung der 95%-Konfidenzintervalle für die Sensitivität eine geringere Effektivität als das konventionelle Screeningverfahren mit manueller Auswertung des Papanicolaou-Tests aufwies und dabei aber höhere Kosten verursachte.

Diskussion

In einem interdisziplinären Netzwerk wurde ein entscheidungsanalytisches Modell zur Evaluation von verschiedenen Screeningprogrammen für das Zervixkarzinom im Kontext des deutschen Gesundheitssystems entwickelt. Dabei wurden deutsche epidemiologische Daten aus verschiedenen Krebsregistern, die aktuelle deutsche Praxis in der Diagnose und Behandlung des Zervixkarzinoms und seiner Vorstufen, sowie deutsche Kostendaten berücksichtigt. Dieses Modell konnte für verschiedene Forschungsfragen bezüglich der Evaluation von medizinischen und ökonomischen Langzeit-Konsequenzen eingesetzt werden.

In der Basisfall-Analyse der Modellrechnung erhöhte ein Screening mit konventioneller Papanicolaou-Methode die Lebenserwartung um 94 Tage und verhinderte 2.994 Krebsfälle pro 100.000 Frauen. Eine Kombination der Dünnschichtpräparation mit anschließender computerisierter Auswertung erhöhte die Lebenserwartung um einen weiteren halben Tag und verhinderte 3.029 Krebsfälle pro 100.000 Frauen gegenüber keinem Screening. Die anderen Screeningstrategien erzielten Ergebnisse, die zwischen dem konventionellen Verfahren und der Kombination der neuen Verfahren lagen. Das konventionelle Screening erwies sich im Vergleich zu der Strategie kein Screening als kosteneffektiv mit einem gesellschaftlich akzeptablen Kosten-Effektivitäts-Verhältnis von 6.600 €/LJ. Der Einsatz neuer Screeningverfahren im Vergleich zum konventionellen Verfahren war mit inkrementellen Kosten-Effektivitäts-Verhältnissen zwischen 220.000 €/LJ und 1.083.000 €/LJ bei jährlichem Screening als nicht kosteneffektiv zu bewerten.

Eine Erhöhung des Screeningintervalls hatte einerseits eine absolute Reduktion in der Effektivität sowohl des konventionellen Screenings als auch bei den neuen

Verfahren, andererseits eine höhere inkrementelle Effektivität der neuen Screeningverfahren im Vergleich zum konventionellen Verfahren zur Folge. Bei einer gesellschaftlichen Zahlungsbereitschaft von 50.000 €/LJ wäre das computergestützte Auswertungsverfahren bei einem vorgegebenen 2-Jahres-Screeningintervall als kosteneffektiv einzuschätzen, die Einführung der Dünnschichtpräparation mit manueller oder computergestützter Auswertung wäre ab einem vorgegebenen Screeningintervall von 3 Jahren als kosteneffektiv einzuschätzen. Diese Ergebnisse werden durch internationale Studien zur Kosten-Effektivität neuer Screeningverfahren gestützt. Die neuen Technologien könnten selbst im Screening mit 2-Jahresintervall eine ähnliche Effektivität wie das konventionelle jährliche Screening erbringen, wobei zum Teil gleichzeitig eine Kosteneinsparung möglich wäre.

Basierend auf diesen Analyseergebnissen ist ein Einsatz der neuen Screeningverfahren anstelle des konventionellen Screenings bei dem aktuell geltenden gesetzlichen Krebsfrüherkennungsprogramms mit jährlichem Screening als nicht kosteneffektiv zu bewerten. Es sollte jedoch kritisch in Betracht gezogen werden, ob ein Einsatz der neuen Screeningmethoden in Verbindung mit einer Erhöhung des Screeningintervalls angestrebt werden sollte. Dabei sind jedoch unbedingt die möglichen Auswirkungen eines solchen Vorgehens auf das Screening-Teilnahmeverhalten zu berücksichtigen.

Die vorliegende klinische und gesundheitsökonomische Entscheidungsanalyse besitzt verschiedene Limitationen, die ihrerseits einen weiteren Forschungsbedarf bestimmen.

Das Modell basiert auf strukturellen Annahmen und numerischen Modellparametern und liefert valide Ergebnisse nur dann, wenn diese Annahmen und Parameter die Wirklichkeit gut widerspiegeln. Es lagen für den deutschen Kontext keine detaillierten Progressions- und Regressionsraten für den natürlichen Verlauf des Zervixkarzinoms vor. Diese Parameter mussten deshalb aus internationalen Studien entnommen und an epidemiologische Beobachtungsdaten von deutschen Krebsregistern angepasst werden. Die Modellergebnisse wurde einer internen und externen Validierung unterzogen, die zeigten, dass das entscheidungsanalytische Modell die epidemiologische Realität in Deutschland valide widerspiegelt.

Bezüglich der Werte für die Sensitivität und Spezifität der einzelnen Screeningtests besteht sowohl große Unsicherheit als auch vermutlich eine große Heterogenität, was die Qualität der Abstriche und ihre Auswertungen betrifft. Viele der Studien zur Bestimmung der diagnostischen Güte von Screeningtests beziehen entweder Risikopopulationen mit erhöhtem a-priori-Risiko für Zervixkarzinom ein, berichten die Ergebnisse nicht stratifiziert für Krankheitsstadien und unterschiedliche Testschwellen-

werte, führen keinen Vergleich mit Goldstandards durch oder unterliegen einem Verification Bias. Für den konventionellen Papanicolaou-Test mit manueller Auswertung konnte auf ein publiziertes Review zurückgegriffen werden, in dem all diese Probleme berücksichtigt wurden und die Daten disaggregiert und um Verification-Bias bereinigt dargestellt waren, so dass hierzu valide gepoolte Schätzer ermittelt werden konnten. Bezüglich der neuen Testverfahren allerdings gibt es noch wenige qualitativ hochwertige Studien, die diese Kriterien erfüllen. Im Rahmen dieser Analyse wurden deshalb ausschließlich relative Testgütwerte in Form von gepoolten Odds Ratios aus Vergleichen zwischen konventionellem und den neuen Screeningtests ermittelt und herangezogen, um die Wahrscheinlichkeit für einen Bias zu minimieren.

Unsicherheit und weiterer Forschungsbedarf besteht bezüglich der exakten Kosten der verschiedenen Screeningtests. Die vorliegende Analyse basiert auf den Abrechnungsgebühren für die Untersuchung im Rahmen des gesetzlichen Früherkennungsprogramms und die zytologische Auswertung des Papanicolaou-Tests. Für die neuen Screeningverfahren wurden über einen Microcosting-Ansatz lediglich die zusätzlichen Materialkosten berücksichtigt. Ein potentieller Mehraufwand oder Einsparungen bezüglich der Personalkosten wurden nicht berücksichtigt.

Ferner besteht Forschungsbedarf hinsichtlich der Lebensqualität von Frauen in den verschiedenen Stadien der Zervixkrankung und des Lebensqualitätseffekts der Durchführung von Konisationen und anderen Behandlungen im Rahmen der Krebsverhütung. Ferner sollte die psychologische Auswirkung positiver Screening-Testbefunde erhoben und in Zukunft in entscheidungsanalytischen Modellen berücksichtigt werden.

Aufgrund fehlender Daten zum Teilnahmeverhalten im deutschen Kontext wurde im vorliegenden Modell die Teilnehmerate vereinfacht als zufällige Teilnahmewahrscheinlichkeit bei jeder Screeninguntersuchung angenommen, die unabhängig von der Geschichte früherer Ereignisse oder Screeningteilnahmen war. Die Teilnehmerate am Screening hatte in Sensitivitätsanalysen einen starken Effekt auf das inkrementelle Kosten-Effektivitäts-Verhältnis der neuen Screening-Verfahren im Vergleich zum konventionellen Verfahren. Es besteht demnach noch Forschungsbedarf einerseits in den Bereichen Psychologie und Public Health, um weitere Evidenz zum Teilnahmeverhalten zu erhalten, und andererseits sind weitere Modellanalysen durchzuführen, in denen die Auswirkungen verschiedener Teilnahmeverhaltensweisen ermittelt wird.

Schließlich ist darauf hinzuweisen, dass diese entscheidungsanalytische Studie ohne die Daten, die den Autoren dieses HTA-Berichts von den Krebsregistern zur Verfügung gestellt wurden, nicht hätte durchgeführt werden können. Krebsregisterdaten wurden in vierfacher Weise in der Modellierung verwendet: (1) zur Kalibrierung des Modells für den deutschen epidemiologischen Kontext, (2) zur Ermittlung der krankheitsspezifischen Mortalität, (3) zur internen Modellvalidierung und (4) zur externen Modellvalidierung. Diese teils getrennt geführten Register werden auch für weitere epidemiologische Analysen und populationsbasierten Hochrechnungen und insbesondere Validierungen von hohem Wert sein.

Schlussfolgerungen

Basierend auf eigenen Modellanalysen für das deutsche Gesundheitssystem können neue Zervixkarzinom-Screeningverfahren zu einer akzeptablen Kosten-Effektivität führen, wenn das Screeningintervall 2 Jahre oder mehr beträgt und eine regelmäßige Teilnahme am Screening gewährleistet ist. Für den derzeitigen deutschen Standard des jährlichen Screenings wird die Einführung der neuen Screeningverfahren als nicht kosteneffektiv bewertet.

Unter der Gewährleistung einer ausreichenden Beteiligungsrate der Frauen am Screeningprogramm sollte jedoch kritisch in Betracht gezogen werden, ob ein Einsatz der neuen Screeningmethoden in Verbindung mit einer Verlängerung des Screeningintervalls auf 2 Jahre angestrebt werden sollte. Dabei sollten jedoch unbedingt die möglichen Auswirkungen eines solchen Vorgehens auf das Screening-Teilnahmeverhalten berücksichtigt werden.

Forschungsbedarf besteht hinsichtlich der validen Testgüte konventioneller und neuer Screeningverfahren unter Routinebedingungen in Deutschland. Ferner sollte das Screening-Teilnahmeverhalten in Deutschland – insbesondere im Hinblick auf eine mögliche Verlängerung des Screeningintervalls – untersucht werden und eine Lebensqualitätserhebung bei Frauen in Abhängigkeit von den Stadien der Zervixkrankung und den Testergebnissen durchgeführt werden.

I Medizinische Effektivität

Verfasser:

Christiane Muth¹, Gaby Sroczynski^{2,3}, Marcial Velasco-Garrido⁴

¹ Abteilung Epidemiologie, Sozialmedizin und Gesundheitssystemforschung, Medizinische Hochschule Hannover, Hannover

² Bayerische Forschungs- und Koordinierungsstelle Public Health, Institut für Medizinische Informatik, Biometrie und Epidemiologie, Ludwig-Maximilians-Universität, München

³ Harvard Center for Risk Analysis, Harvard School of Public Health, Boston, USA

⁴ Institut für Gesundheitswissenschaften, FG Management im Gesundheitswesen, Technische Universität Berlin

I C Hauptdokument Medizinische Effektivität

Verfasser:

Christiane Muth, Marcial Velasco-Garrido

I C 1 Policy Question

Wie in anderen westlichen Industrienationen ist auch in Deutschland mit Einführung des Krebsvorsorgeprogramms bei Frauen ein deutlicher Rückgang der Zahl der Neuerkrankungen wie auch in der Sterblichkeit an Gebärmutterhalskrebs erkennbar (Levi et al. 2000). Zählt dieser Tumor weltweit noch zu den häufigsten Krebserkrankungen der Frau, insbesondere in Ländern ohne Früherkennungsmaßnahmen, wie z.B. in Zentral- und Südamerika oder in der Sub-Sahara-Zone Afrikas (Boyle et al. 2000), ist er in Deutschland inzwischen an die achte Stelle gerückt (Robert-Koch-Institut 2000). Da ein bundesweites Krebsregister fehlt, wird die Neuerkrankungsrate (Inzidenz) aus den Angaben der Register im Saarland und in den neuen Bundesländern geschätzt. In Deutschland erkranken danach jährlich etwa 6200 Frauen an der invasiven Form des Gebärmutterhalskrebses. Die standardisierten Inzidenzraten fielen seit den siebziger Jahren bis zum Beginn der achtziger Jahre auf weniger als die Hälfte der Ausgangswerte, danach gingen sie nur noch geringfügig zurück (Gesundheitsbericht für Deutschland 1998). 1995 verstarben in Deutschland 2.207 Frauen an einem Zervixkarzinom; die standardisierte Mortalitätsrate lag bei 4,2 je 100.000 Frauen. Seit den siebziger Jahren geht die Mortalität deutlich zurück, in den letzten 15 Jahren allein um rund 40% (Gesundheitsbericht für Deutschland 1998).

Seit 1971 sind in Deutschland Krebsfrüherkennungsuntersuchungen in den Leistungskatalog der gesetzlichen Krankenversicherung (GKV) einbezogen und als Krebsfrüherkennungsprogramm bundesweit eingeführt. Nach den gesetzlichen Grundlagen (§25 Abs. 2 SGB V) haben Frauen ab dem 20. Lebensjahr einmal jährlich Anspruch auf eine Früherkennungsuntersuchung auf Zervixkarzinome.

Die Früherkennung des Zervixkarzinoms erfolgt durch einen Abstrich im Rahmen einer gynäkologischen Untersuchung, durch den sogenannten Papanicolaou-Test (im weiteren kurz als Pap Test bezeichnet). Der Abstrich hat zum Ziel, in einer zytologischen Begutachtung Krebsvorstufen im Sinne einer Sekundärprävention zu erkennen. Untersuchungen haben gezeigt, dass die Fehlerrate des zytologischen Tests beträchtlich sein kann und sich durch Qualitätskontrollen verbessern lässt. Weitere Ansätze bilden neue Präparationstechniken und computergestützte Auswertungs-

technologien, die zum einen die Qualität des Zellabstriches verbessern können, zum anderen falsch-negative Befundraten verringern sollen.

Nach den bestehenden gesetzlichen Grundlagen (§2 Abs. 1 SGB V) haben Qualität und Wirksamkeit der erbrachten Leistungen im Gesundheitswesen dem allgemein anerkannten Stand der medizinischen Erkenntnisse zu entsprechen und den medizinischen Fortschritt zu berücksichtigen.

Im Rahmen des Wirtschaftlichkeitsgebots (§12 Abs. 1 SGB V) dürfen die Leistungen das Maß des Notwendigen nicht überschreiten. Neue Untersuchungs- und Behandlungsmethoden in der ambulanten Vertrags(zahn)ärztlichen Versorgung sind nur dann abrechnungsfähig, wenn der Bundesausschuss der Ärzte und Krankenkassen deren diagnostischen oder therapeutischen Nutzen sowie deren medizinische Notwendigkeit und Wirtschaftlichkeit - auch im Vergleich zu bereits zu Lasten der Krankenkassen erbrachten Methoden - anerkannt hat (§135 Abs.1 SGB V).

Angesichts dieser gesetzlichen Bestimmungen sollte vor der Entscheidung über die Aufnahme neuer zytologischer Technologien in den Leistungskatalog der gesetzlichen Krankenversicherung eine Bewertung ihrer klinischen Effektivität und eine Abschätzung der ökonomischen Konsequenzen erfolgen. Anhand der zum heutigen Zeitpunkt verfügbaren Evidenz sollte daher geklärt werden, inwieweit sich durch den Einsatz dieser Technologien die Effektivität des Zervixkarzinom-Screenings steigern lässt und mit welchen Auswirkungen ihrer Einführung auf die Kosten und Kosten-Effektivität des Zervixkarzinom-Screenings gerechnet werden muss. Insgesamt soll damit ein Beitrag zur sinnvollen Ressourcenallokation im Gesundheitswesen geleistet werden.

Bleiben diese Technologien aus dem Leistungskatalog der GKV ausgeschlossen, werden sie weiterhin als „IGEL-Leistungen“ (die private Inanspruchnahme individueller Gesundheitsleistungen für gesetzlich krankenversicherte Patientinnen und Patienten) gegen Gebühr angeboten. In diesem Kontext soll dieser HTA-Bericht auch einen Beitrag unter Verbraucherschutzaspekten leisten. Frauen, die sich zum Zervixkarzinom-Screening entschließen, sollten das auf der Basis einer informierten Zustimmung tun. Auf der Grundlage der in der vorliegenden Arbeit ermittelten Evidenzlage erstellte Patienteninformationen können eine beratende Funktion bei der Entscheidungsfindung der Frauen leisten, ob und wenn ja, zu welchen der auf dem Markt angebotenen Vorsorgemaßnahmen gegen Gebärmutterhalskrebs sie zustimmen. Insbesondere können sie befähigt werden, von Herstellerinteressen unabhängige Entscheidungen zu treffen, wenn es um finanzielle Mehraufwendungen für sie geht.

Der Bericht wurde im Auftrag des Bundesgesundheitsministeriums im Rahmen des Projekts „Weiterentwicklung einer Datenbasis ‘Evaluation medizinischer Verfahren und Technologien (Health Technology Assessment)’ in der Bundesrepublik Deutschland“ erstellt.

I C 2 Hintergrund

Verfasser:

Gaby Sroczynski, Christiane Muth

I C 2. 1 Beschreibung des Krankheitsbildes

I C 2. 1. 1 Definition des Zervixkarzinoms

Das Zervixkarzinom (Gebärmutterhalskrebs, Synonyme: Portiokarzinom, Kollumkarzinom) ist ein bösartiger Tumor, der vom unteren Drittel des Uterus (Gebärmutter) ausgeht. In Frühstadien ist er meistens symptomfrei, später können Blutungsunregelmäßigkeiten auftreten, in Spätstadien Schmerzen und Störungen der Blasen- und Mastdarmfunktion, die auf einer kontinuierlichen Tumorausbreitung in die angrenzenden Organe im Beckenbereich beruhen. Fernmetastasen in Lymphknoten, die im Bereich von Bauchschlagader und Mediastinum (Mediastinum: "Mittelfell", Bereich des Brustkorbs, in dem u.a. Herz, große Gefäße, Luft- und Speiseröhre liegen, der nach unten vom Zwerchfell begrenzt wird und der nach oben in den Hals übergeht. (David, 1987)) angesiedelt sind, sowie in Lunge und Skelett treten erst relativ spät auf. (AJCC 1997; Zollinger 1981, Gesundheitsberichterstattung Kapitel 5.7 1998). Verschlüsselt wird der Tumor nach ICD-10 mit „C53“; nach ICD-9 mit „180“.

Die Krebsentstehung beginnt mit Fehlentwicklungen der Gebärmutter Schleimhaut (Zervixdysplasien; Cervical Intraepithelial Neoplasia), die unbehandelt in ein Carcinoma in situ (Oberflächenkarzinom) und schließlich in ein invasives Zervixkarzinom übergehen können (Zollinger 1981).

I C 2. 1. 2 Epidemiologie des Zervixkarzinoms

Das Zervixkarzinom ist weltweit das dritthäufigste Karzinom der Frau (Boyle et al. 2000), in Deutschland liegt es, bezogen auf die Neuerkrankungsfälle, an achter Stelle (Robert-Koch-Institut 1999). Achtzig Prozent der weltweit etwa 500.000 neuen Fälle pro Jahr treten in den Entwicklungsländern auf. Obwohl das Zervixkarzinom in diesen Ländern die häufigste bösartige Erkrankung der Frau darstellt, wurden in den ver-

gangenen fünf Jahren lediglich 5% der weiblichen Bevölkerung auf Gebärmutterhalskrebsvorstufen hin untersucht (Spitzer 1998). In den USA und den meisten Ländern Europas ist in den letzten 20 Jahren ein kontinuierlicher Rückgang der Inzidenzraten des Zervixkarzinoms zu beobachten (Baltzer et al. 2000), welches unter anderem auf ein effektives Zervixkarzinom-Screening zurückzuführen ist (Hillemanns et al. 1997). Spiegelbildlich kam es in diesem Zeitraum zu einem Anstieg präinvasiver Neoplasien der Zervix (Köchli et al. 1998), ein Phänomen der Einführung von Screeningmaßnahmen.

Abbildung 1 zeigt die Entwicklung der Teilnehmerate am Zervixkarzinom-Screening in den USA bei gleichzeitiger Darstellung von Inzidenz und Mortalität des invasiven Zervixkarzinoms im zeitlichen Verlauf (Kurman 2002).

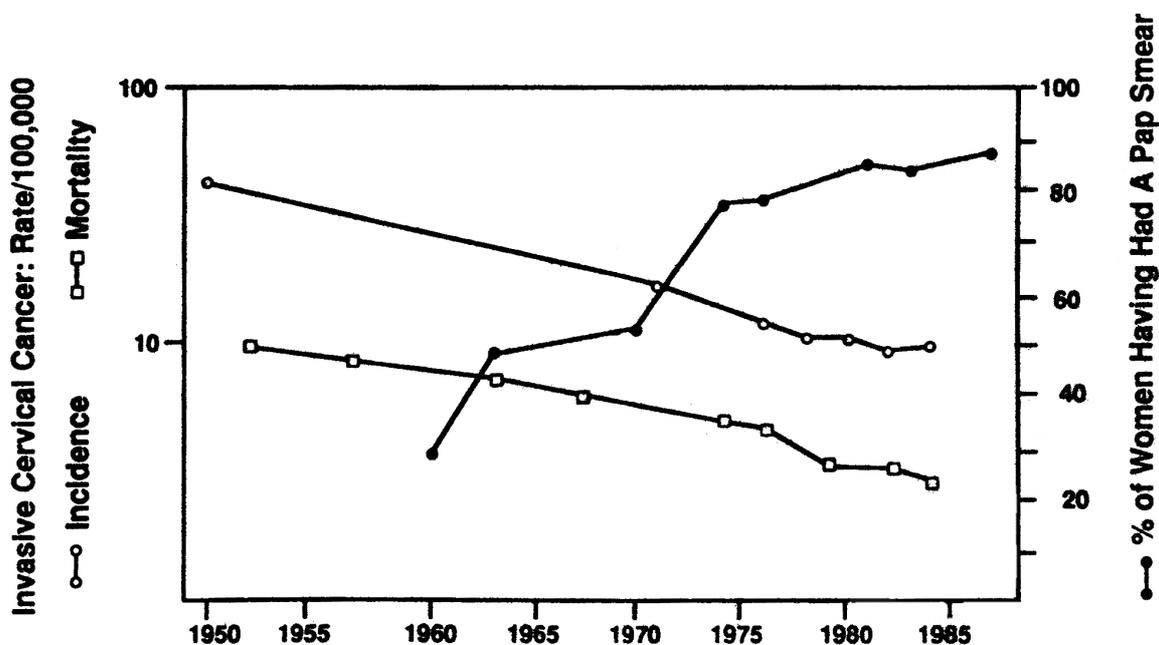


Abbildung 1: Screeningteilnahmeraten, Inzidenz und Mortalität des Zervixkarzinoms in den USA

Quelle: Kurman; 2002.

Das Lebenszeitrisiko einer Frau, an Zervixkarzinom zu erkranken, wird in den USA mit 0,81% angegeben (SEER-Programm 2002).

In Deutschland erkranken derzeit etwa 7000 Frauen jährlich an Zervixkarzinom (Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland 2002). Das entspricht einem Anteil von 3 - 4% an allen bösartigen Neubildungen bei Frauen, der jedoch altersabhängig stark variiert. Bei jeder vierten an Krebs erkrankten Frau

im Alter von 25-35 Jahren wird Gebärmutterhalskrebs diagnostiziert, demgegenüber liegt dieser Anteil bei den über 65-Jährigen unter 5%. Das mittlere Erkrankungsalter liegt bei 55,1 Jahren. Das Zervixkarzinom zeigt in seiner Inzidenz einen ersten Altersgipfel zwischen dem 40. und 50. Lebensjahr und einen zweiten zwischen dem 65. und 75. Lebensjahr (Abbildung 2).

Die Zervixkarzinominzidenz liegt im europäischen Vergleich im oberen Drittel, wobei innerhalb Deutschlands erhebliche regionale Unterschiede zu vermerken sind. In Portugal, Dänemark und Österreich liegen ebenfalls hohe Erkrankungsraten vor, während in Spanien, Finnland und den Niederlanden niedrige Inzidenzen berichtet werden (Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland, 2002). Die höheren Karzinominzidenzen in Deutschland, Dänemark und Österreich sind in höheren Ausgangsraten begründet. Von Gustafson et al. (1997) liegen Berechnungen der altersstandardisierten Inzidenzraten vor Einführung des Screenings vor: für Deutschland betrug sie 35,8 , für Dänemark 31,3 , für die Niederlande 18,9 und für Finnland 14,8 je 100.000 Frauen.

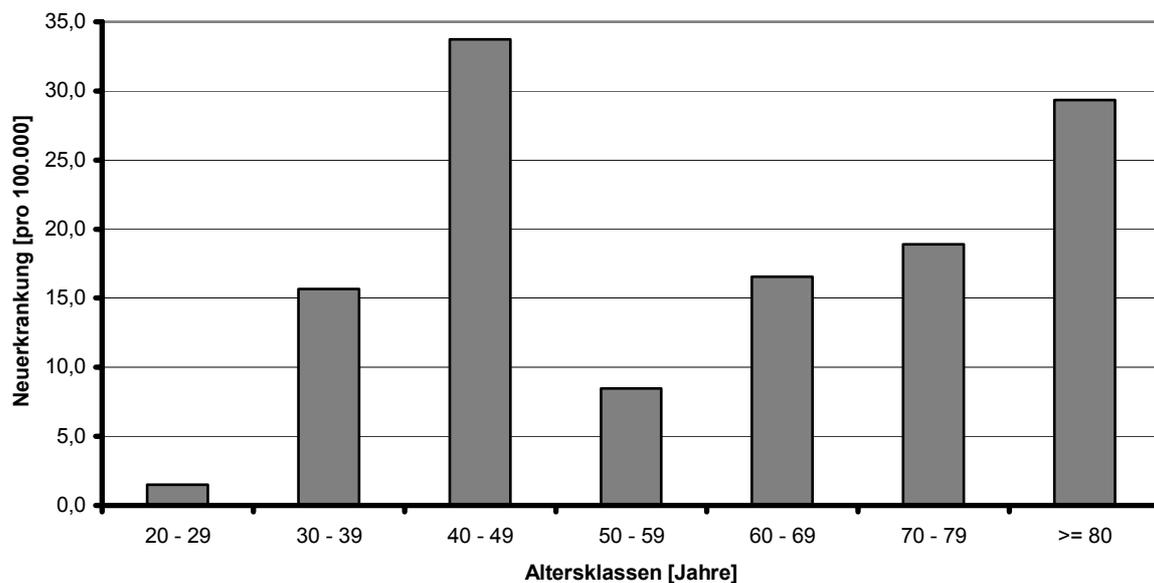


Abbildung 2: Altersspezifische Inzidenz des Zervixkarzinoms im Saarland

Quelle: Daten des Saarländischen Krebsregisters. Altersspezifische Inzidenzdaten des Jahres 1996.

Im zeitlichen Verlauf weist die Inzidenz des Zervixkarzinoms in Deutschland wie auch in Europa bis Mitte der achtziger Jahre einen deutlichen Rückgang auf und ist seit-

dem annähernd konstant geblieben (Abbildung 3) (Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland 2002).

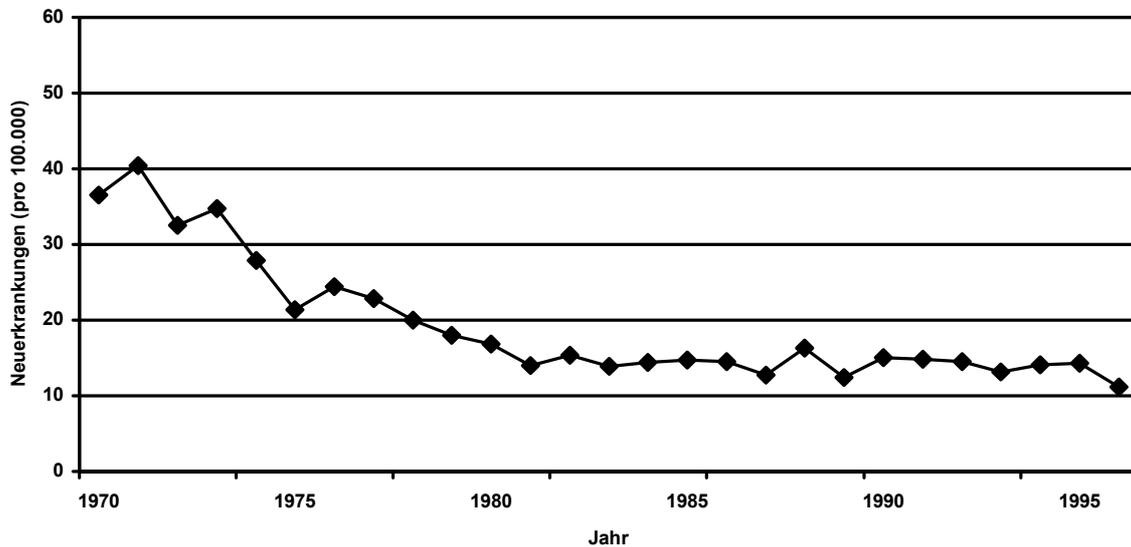


Abbildung 3: Zeitlicher Verlauf der Zervixkarzinominzidenz in Deutschland

Indikator: Altersstandardisierte Erkrankungsrate, alle Altersgruppen, standardisiert auf neue Europa-standardbevölkerung, basierend auf den Daten des saarländischen Krebsregisters. (Quelle: Robert Koch-Institut. Daten des saarländischen Krebsregisters von 1970-1996)

Betrachtungen über die Mortalität des Zervixkarzinoms können unter verschiedenen Gesichtspunkten angestellt werden und lassen sich mit verschiedenen Maßzahlen beschreiben. Die Deskription der zeitlichen Entwicklung der Mortalität über alle Altersgruppen erfolgt mit altersstandardisierten Raten. In Deutschland ist die altersstandardisierte Mortalitätsrate seit 1970 um 55% gesunken (Abbildung 4) (Daten des Statistischen Bundesamts BRD; Tumorregister München, Jahresbericht 1999).

Werden die altersspezifischen Mortalitätsraten, d.h. die auf bestimmte Altersgruppen eingeschränkten Mortalitätsraten der Jahre 1980 bis 1997 verglichen, ist eine Abnahme erst ab dem Alter von 50 Jahren erkennbar.

Auch weist die Mortalität beim Zervixkarzinom deutliche regionale Unterschiede auf. So liegt sie z.B. in den neuen Bundesländern ca. 60% über der Mortalität der alten Bundesländer, die Region München weist eine 40% niedrigere Mortalität im Vergleich zum Bundesdurchschnitt auf (Engel & Hölzel 2001).

Alle Angaben zur Mortalität weisen Unsicherheiten auf, die in der Ungenauigkeit der Todesursachenstatistik bzw. in der unvollständigen Erfassung der einzelnen Register begründet sind, da kein bundesweites zentrales Tumorregister existiert.

Der Verlust an Lebenserwartung für eine an Gebärmutterhalskrebs erkrankte Frau liegt im Durchschnitt bei 9 Jahren (Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland 1999).

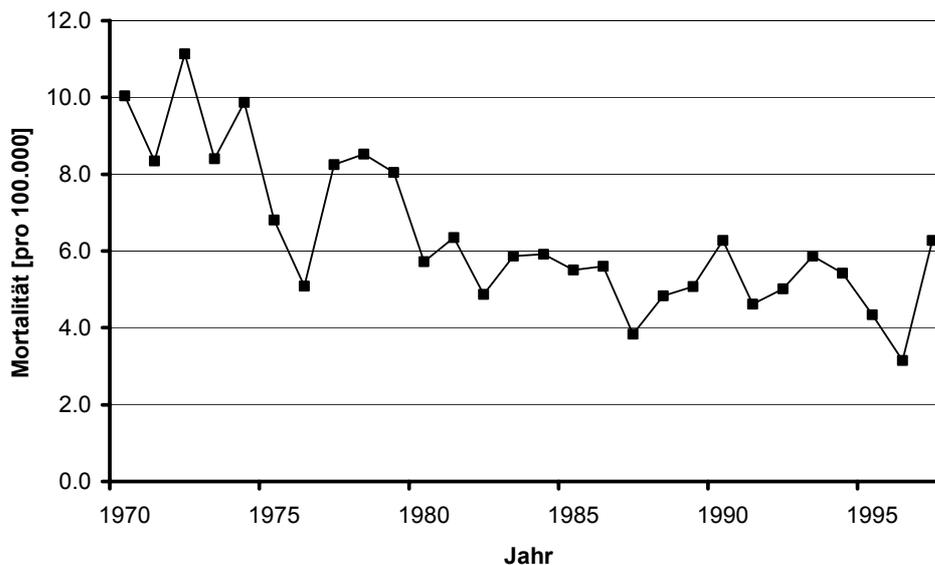


Abbildung 4: Altersstandardisierte Zervixkarzinom-Mortalitätsrate in Deutschland

Quelle: Robert-Koch-Institut. Daten des Saarländischen Krebsregisters

Die relativen 5-Jahres- und 10-Jahres-Überlebensraten beim Zervixkarzinom liegen bei 66% und 59% (Engel et al. 2000), siehe auch Abbildung 6 in Kapitel I C 2. 1. 7 Im Trend zeigt sich seit 1988 eine signifikante Verbesserung des 5- und 10-Jahres-Überlebens. Stadienspezifisch ist im Trend jedoch nur ein geringer Überlebensvorteil zu erkennen, so dass eine Verbesserung des Gesamtüberlebens auf einen Wechsel in den Therapieformen und auf eine Verschiebung zu prognostisch günstigeren Karzinomstadien zurückzuführen ist. Bei deutlicher Zunahme der Stadien (für eine ausführliche Beschreibung der Stadien siehe Kapitel A.2.1.4, insbesondere Tabelle 4) IA und IB konnte gleichzeitig ein Rückgang in den Stadien IIA und v.a. IIB beobachtet werden, sowie eine leichte Zunahme des Stadium IV (Engel & Hölzel 2001).

I C 2. 1. 3 Ätiologie und Pathologie des Zervixkarzinoms

Ungefähr 85% - 90% der Zervixkarzinome sind Plattenepithelkarzinome (Riethdorf; 1999b). Die übrigen 10% - 15% sind Adenokarzinome, ausgehend vom Drüsenepithel, adenosquamöse Karzinome (gemischte Karzinome) und andere seltene Tumorarten. Allen diesen selteneren Tumorarten ist gemeinsam, dass sie in einem Papanicolaou Testverfahren (Pap Test) schlechter zu diagnostizieren sind (DeMay; 1997) und dass sie stadienunabhängig zumeist eine schlechtere Prognose aufweisen (Canavan 2000). In der vorliegenden Arbeit geht es nur um Plattenepithelkarzinome.

Ätiologie und Risikofaktoren

Als Hauptrisikofaktor wird in der Literatur eine persistente Infektion mit dem Humanen Papillomavirus (HPV) diskutiert (Baltzer et al. 2000). Die HPV-Infektion der Hochrisiko-HPV-Typen 16, 18, 31, 33, 45, 51, 52 und 56 (Das Humane Papillomvirus lässt sich mit molekularbiologisch-virologischen Methoden in zahlreiche Typen unterscheiden. Bisher wurden auf diese Art und Weise mehr als 100 Typen isoliert, von denen jedoch nicht alle ein onkogenes, d.h. krebserregendes Potential haben.) gilt nach neuesten Erkenntnissen als Hauptursache für die Entstehung der Vorstufen des Zervixkarzinoms (Schneider et al. 1999). Mit modernen Testmethoden konnte die HPV-DNA bei mehr als 95% der Zervixkarzinompatientinnen nachgewiesen werden (Schiffman et al. 1993 in: Cuzick et al. 1999). Frauen, die mit onkogenen HPV-Typen infiziert wurden, haben ein 40 – 180-fach erhöhtes Risiko, bei Persistieren der HPV-Infektionen sogar ein 100 – 500-fach erhöhtes Risiko, hochgradige Zervixepithelveränderungen zu entwickeln (IARC 1995 und Olsen et al. 1995 in: Cuzick et al. 1999).

Die Prävalenz von HPV liegt bei schwergradigen Dysplasie und beim invasiven Zervixkarzinom zwischen 75% und 95%; bei gesunden Frauen liegt die Prävalenz bei 3-20% (Cuzick et al. 1999). Für Deutschland bleibt die HPV-Prävalenz bislang unbekannt. Eine Studie (Koch et al. 1997) in im Stadtgebiet Berlin berichtete eine HPV-Prävalenz von 19,7%, während eine andere Studie (Schneider et al. 2000) aus Thüringen eine HPV-Prävalenz von 7,8% berichtete.

Vermutet wird ein onkogenes Potential auch bei anderen Mikroorganismen. Anttila et al. (2001) fanden bei Patientinnen mit Zervixkarzinom eine Assoziation zu Infektionen mit bestimmten Serotypen von *Chlamydia trachomatis*, die Datenlage ist hierzu jedoch unsicher.

Bereits vor Nachweis des Kausalzusammenhangs zwischen Zervixkarzinom und HPV-Infektion waren in zahlreichen Untersuchungen Risikofaktoren aus dem Bereich

des Sexualverhaltens ermittelt worden. So sind insbesondere frühe Aufnahme des Sexualverkehrs und häufig wechselnde Geschlechtspartner mit der Erkrankung assoziiert. Niedriger Sozialstatus, Langzeittherapie mit oralen Kontrazeptiva und Rauchen sind ebenfalls mit einem erhöhten Risiko vergesellschaftet (Boyle et al. 2000).

Die Erkenntnis, dass persistente HPV-Infektionen den Hauptrisikofaktor in der zervikalen Karzinogenese bilden, hat das Verständnis der epidemiologischen Forschung im Bereich Zervixkarzinom grundlegend geändert. Alle bisher untersuchten Risikofaktoren müssen nun auf ihre Unabhängigkeit bezüglich des HPV untersucht werden (Schneider et al. 1999).

Pathologie

Der Gebärmutterhals, die Zervix uteri, wird in einen außen liegenden, zapfenförmig in die Scheide ragenden Teil (die Ektozervix oder Portio vaginalis) und einen innen den Gebärmutterkanal umschließenden Teil (die Endozervix oder Portio supravaginalis) eingeteilt. An der Grenze zwischen beiden Teilen befindet sich im Schleimhautniveau die sogenannte Transformationszone, da dort verschiedene Arten von Schleimhaut aneinandergrenzen: das außen liegende Plattenepithel und das innen liegende Zylinderepithel. Diese Transformationszone ist die häufigste Lokalisation für neoplastische Veränderungen (Cannistra & Niloff 1996). Die anatomische Lage dieser Umwandlungszone ist individuell sehr unterschiedlich und hängt u.a. von Alter, Form und Lage der Gebärmutter und dem Geburtenstatus ab. Grundsätzlich liegt sie bei jungen Frauen eher im gut zugänglichen Ektozervixbereich, während sie mit zunehmendem Alter eher endozervikal und damit schlechter zugänglich lokalisiert ist.

Früh- und Vorstadien des Zervixkarzinoms sind aus diesem Grund bei jungen Frauen vorwiegend an der Portiooberfläche, bei älteren und alten Frauen häufiger endozervikal lokalisiert (Baltzer et al. 2000).

Das Zervixkarzinom entwickelt sich in den meisten Fällen langsam über einen Zeitraum von mehreren Jahren aus seinen Vorstufen (Abbildung 5), den zervikalen intraepithelialen Neoplasien (CIN) und dem Carcinoma in situ (CIS). Die Prävalenz dieser Vorstufen zeigte in Westeuropa und den USA eine ständige Zunahme in den letzten zwei Dekaden und liegt bei ca. 3-5%. Frauen im reproduktiven Alter zwischen 25 und 40 Jahren sind vorwiegend betroffen. Bezüglich der Latenzzeiten zur Entwicklung eines invasiven Karzinoms liegen keine Follow-up-Studien vor, aufgrund der guten Behandelbarkeit von Vorstufen wären derartige Beobachtungsstudien zum natürlichen Verlauf auch nicht gerechtfertigt. In der Literatur werden Latenzzeiten von zehn Jahren und mehr (Barron & Richart 1968, 1970) genannt, die in statistischen

Modellrechnungen ermittelt wurden. Plaxe & Saltzstein (1999) berechneten aus den Daten des SEER-Programms sogar bis zu 18 Jahre Latenzzeit, wiesen jedoch darauf hin, dass alle Daten, die in einer solchen Art und Weise gewonnen werden, einem *Ecological Fallacy* (Zusammenhänge, die zwischen aggregierten Variablen beobachtet werden, die jedoch nicht notwendigerweise die Zusammenhänge repräsentieren, die auf individueller Ebene existieren (Gordis 2001). Ecological Fallacy ist ein typisches und schlecht zu kontrollierendes Problem in ökologischen Studien, wie in der von Plaxe & Saltzstein, in der mit aggregierten Daten aus Bevölkerungsregistern gearbeitet wurde.) unterliegen können.

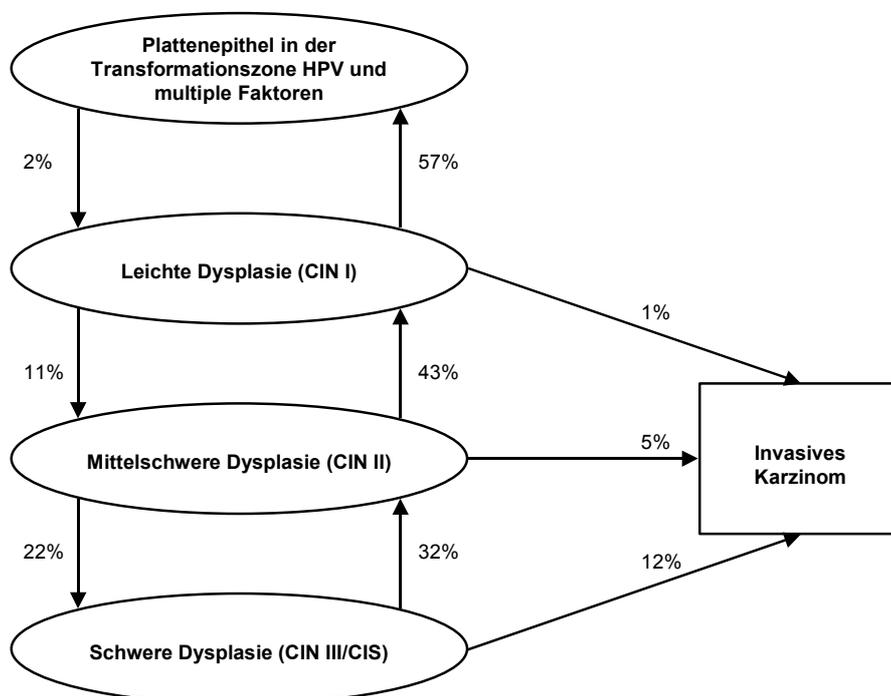


Abbildung 5: Karzinogenese der Zervix uteri

Modifiziert nach Köchli et al. (1998), Progressions-, Regressions- und Invasionsdaten aus Östör (1993) und Sherlaw-Johnson (1994); CIN= Cervical Intraepithelial Neoplasia, CIS= Carcinoma in situ.

Unbehandelte leichte Dysplasien der Portio uteri (CIN I) gehen in ungefähr 11% der Fälle in ein Carcinoma in situ und zu 1% in ein invasives Zervixkarzinom über. Die mittlere Regressionsrate liegt bei 57% (Östör 1993). Selbst bei der CIN II ist in 43% der Fälle mit einer spontanen Regression der Läsion zu rechnen, allerdings auch in 22% der Fälle mit einer Progression zum CIN III / CIS und in 5% zu einem invasiven Karzinom. Die hohe Rate an spontanen Regressionen leichtgradiger Zervixdysplasien beruht auf der Tatsache, dass dysplastisch veränderte Schleimhautepithelien

Folge von HPV-Infektionen sind, die aufgrund der immunologischen Abwehr spontanen Heilungsvorgängen unterliegen.

I C 2. 1. 4 Nomenklatur

Klassifikation der Zervixdysplasien

Da Nomenklaturen von Erkrankungen aus systematischen Konzepten bestehen, die morphologische Merkmale, Verständnis von Krankheitsentwicklung sowie prognostische Indices beinhalten, sind sie einem laufenden Wandel unterworfen. Trotz eines allgemeinen Trends zu internationaler Vereinheitlichung und einer weitgehenden Übereinkunft bei der Stadieneinteilung des manifesten Zervixkarzinoms bestehen insbesondere in der Klassifikation zytologischer und histologischer Befunde derzeit zahlreiche Nomenklatorsysteme nebeneinander, die untereinander nur bedingt vergleichbar sind. Die wichtigsten werden in den folgenden Abschnitten beschrieben.

Befundwiedergabe in der Zytologie des Zervixabstrichs

In der vorliegenden Arbeit beschränken wir uns auf die Darstellung des im angloamerikanischen Sprachraum wie auch in allen von uns identifizierten Studien eingesetzten Bethesda-Systems in seiner Fassung von 1991 (Anonymus 1992), da die im September 2001 vorgestellte Fassung (Bethesda System 2001; www.bethesda2001.cancer.gov) noch keinen Eingang in die von uns bewerteten Studien gefunden hat und stellen ihr die in Deutschland gebräuchlichste Klassifikation nach der Münchner Nomenklatur II (1991) in der Ergänzung aus dem Jahre 1997 (Schenck 2001) gegenüber. Der bei der Befunddokumentation verwendete sogenannte Vorsorgeschein (siehe Anhang) basiert auf der Münchner Nomenklatur I von 1975. Die letzten Änderungen von 1997, die insbesondere um Angaben zur Qualität des Abstrichs und Empfehlungen zum weiteren klinischen Management erweitert wurden, sind bislang dort nicht eingegangen (Schenck 2001).

Die Befundwiedergabe umfasst neben diagnostischen Aussagen auch Angaben über die Abstrichqualität (*Adequacy*), z.B. ob ausreichend viel Material in technisch einwandfreiem Zustand zur Beurteilung vorliegt (im Bethesda-System als *Satisfactory* bezeichnet), ob es z.B. durch Beimengungen von Blut, Eiter oder Schleim nur eingeschränkt beurteilbar ist (*Satisfactory But Limited by Results; SBLB*) oder ob es nicht beurteilbar ist (*Unsatisfactory*). Begriffsdefinitionen der einzelnen Gruppen der

Münchener Nomenklatur II (nach Schenck 2001) und des Bethesda-Systems (nach Anonymus 1992) sind in den folgenden Tabellen zusammengefasst.

Tabelle 1: Klassifikation zytologischer Befunde in der Münchener Nomenklatur II (Schenck 2001)

Gruppe	Begriffsdefinition	Empfehlung
I	Normales Zellbild, altersentsprechend, einschl. leichter entzündlicher und degenerativer Veränderungen sowie bakterieller Zytolyse	
II	Deutliche entzündliche Veränderungen, Regenerationsepithel, unreife metaplastische Zellen, stärkere degenerative Veränderungen, Para- und Hyperkeratosezellen, spezielle Zellbilder wie z.B. HPV-Infektionen ohne wesentliche Kernveränderungen u.a.	Ggf. zytologische Kontrolle, Zeitabstand je nach klinischem Befund
III D	Zellen einer Dysplasie leichten bis mäßigen Grades; besondere Erwähnung bei Zeichen einer HPV-Infektion	Kolposkopisch-zytologische Kontrolle in drei Monaten
IV a	Zellen einer schweren Dysplasie oder eines Carcinoma in situ; besondere Erwähnung bei HPV-Infektion;	Kolposkopisch-zytologische Kontrolle u. Histologische Klärung;
IV b	Zellen einer schweren Dysplasie oder eines Carcinoma in situ; Zellen eines invasiven Karzinoms nicht auszuschließen	Kolposkopisch-zytologische Kontrolle u. Histologische Klärung
V	Zellen eines malignen Tumors (auch extragenitalen Ursprungs)	Kolposkopisch-zytologische Kontrolle und Histologische Klärung
III	Unklarer Befund (schwere entzündliche, degenerative o. a. Veränderungen, die eine sichere Zuordnung zwischen gut- und bösartigen Befunden nicht zulassen)	(je nach klinischem Befund kurzfristige zytologische Kontrollen oder histologische Abklärung)

Tabelle 2: Epitheldysplasien im Bethesda-System 1991 (Ausschnitt)

	Klassifikation	Beschreibung
Plattenepithel	ASCUS (<i>Atypical Squamous Cells of Undetermined Significance</i>)	Atypische Plattenepithelzellen unklarer klinischer Bedeutung
	LSIL (<i>Low Grade Squamous Intraepithelial Lesion</i>)	Milde Dysplasien, HPV-Infektionen, einschl. CIN I
	HSIL (<i>High Grade Squamous Intraepithelial Lesion</i>)	Mäßige und schwere Dysplasien; einschl. CIN 2/ CIN 3 /CIS
	Squamous Cell Carcinoma	Invasives Plattenepithelkarzinom
(Drüsenepithel)	AGUS (<i>Atypical Glandular Cells of Undetermined Significance</i>)	Atypische Drüsenepithelzellen unklarer klinischer Bedeutung
	(Adenocarcinoma)	Verschiedene Arten von Adenokarzinomen

Die Befundkategorie AGUS ist hier auch dargestellt worden, obwohl es sich dabei nicht um eine Plattenepitheldysplasie handelt. In vielen Studien werden die Kategorien ASCUS und AGUS jedoch nicht separat erfasst, da sich ihre Unterscheidung mitunter schwierig gestaltet und AGUS nur mit geringer Häufigkeit auftritt.

Eine Gegenüberstellung beider Systeme sowie ihr Abgleich zu histologischen Befunden ist in Tabelle 3 dargestellt.

Befundwiedergabe in der Histologie einer Zervixbiopsie / einer Zervixabrasio

Werden bei einer kolposkopischen Untersuchung (Kolposkopie: siehe Kapitel I C 2. 1. 5) der Portio, des Teiles der Gebärmutter, der in die Vagina hineinragt, Gewebsveränderungen dargestellt, können diese biopsiert und histologisch untersucht werden. Bei den Teilen der Gebärmutter, die hinter der Portio im Gebärmutterkanal liegen, die somit einer Probenentnahme nicht oder nur schlecht zugänglich sind, wird zur Probengewinnung eine Abrasio (Ausschabung) eingesetzt.

Wie in der Zytologie existieren für die histologische Befundung mehrere Nomenklaturesysteme im internationalen Vergleich nebeneinander. Die weiteste Verbreitung fand das von Richart vorgeschlagene System zervikaler intraepithelialer Neoplasien verschiedener Schweregrade (Richart 1973 in Broadstock, 2000) CIN 1, CIN 2, CIN 3, wobei CIN 3 auch in-situ-Karzinome einschließt. Später wurde von ihm eine weitere Vereinfachung des Systems in nur zwei Kategorien Lo-CIN (low grade) und Hi-CIN (high grade) vorgestellt (Richart 1990 in Broadstock, 2000). Eine Gegenüberstellung dieser Nomenklaturen und der von der WHO / ISGYP (Reagan 1979 in Broadstock, 2000) vorgeschlagenen Klassifikation mit der zytologischen Befunddarstellung zeigt Tabelle 3.

Gegenüberstellung zytologischer und histologischer Nomenklaturen

In der folgenden Tabelle werden die Nomenklatorsysteme gegenübergestellt, die in Deutschland gebräuchlich sind bzw. die in den von uns ausgewerteten Studien verwendet wurden.

Tabelle 3: Schematische Gegenüberstellung von Nomenklaturen

Histologie	Richart (1973)				Kondylome	Cervical Intraepithelial Neoplasia (CIN)				
	(1990)					CIN 1	CIN 2	CIN 3		
	Reagen (1979) (WHO/ ISGYP (International Society of Gynecological Pathologist))		Atypie		Kondylomartige Läsionen	Leichte Dysplasie	Mäßige Dysplasie	Schwere Dysplasie	Ca in situ (CIS)	
Zytologie	Bethesda System (BS) (1991)	Normal	Infektion	ASCUS	Squamous Intraepithelial Lesion (SIL)					Invasives Karzinom
			Reaktive Veränderungen		Low grade (LSIL) (including HPV)		High grade (HSIL)			
	Papanicolaou	I	II		III				IV	V
	Münchener Nomenklatur II (1997)	I	II		III D		IV a	IV a IV b	V	

Histologische und zytologische Klassifikationssysteme bei Befunden des Zervixepithels. (Modifiziert nach Nanda et al. 2000 in Broadstock et al. 2000).

Wie in der Darstellung deutlich wird, sind die Systeme untereinander nur bedingt vergleichbar, da es sich um ein Kontinuum von Veränderungen handelt, die durch Festlegung unterschiedlicher Schwellenwerte in einzelne, nicht identische Gruppen zusammengefasst wurden.

Klassifikation des Zervixkarzinoms

Das invasive Zervixkarzinom wird international übereinstimmend nach Richtlinien der Fédération Internationale de Gynécologie et d'Obstétrique (FIGO) oder nach der von der Union Internationale Contre le Cancer (UICC) vorgeschlagenen TNM-Klassifikation eingeteilt. Ziel ist es, den Schweregrad der Erkrankung zu bestimmen, der für die weitere Therapieplanung einen wichtigen Baustein darstellt. Im sogenannten klinischen Staging zur Bestimmung des FIGO-Stadiums wird der Ausbreitungsgrad des Tumors zum Zeitpunkt der Diagnosestellung allein klinisch und präoperativ

bestimmt. In die FIGO-Klassifikation fließen die Ergebnisse aus folgenden Untersuchungen ein: Palpation (ggf. in Narkose), Inspektion, Zytologie, Kolposkopie, Biopsie (bzw. Kürettage), Zystoskopie, Rectosigmoidoskopie, intravenöses Urogramm, Röntgenuntersuchung von Lunge und Skelett (Wulf KH & Schmidt-Matthiesen H, 1991). Nach den Vorschlägen der FIGO erfolgt bei Vorliegen weiterer Untersuchungsergebnisses keine Korrektur des ursprünglichen klinischen Tumorstadiums (FIGO 1998, Kurman 2002). Damit soll sichergestellt werden, dass die Stadieneingruppierung für alle betroffenen Patientinnen unabhängig vom Ausstattungsgrad eines Krankenhauses vorgenommen wird.

Üblicherweise werden die klinischen Tumorstadien nach FIGO angegeben, werden sie in der TNM-Klassifikation dargestellt, verwendet man die Kennzeichnung cTxNxMx. Bei der TNM-Nomenklatur handelt es sich um eine allgemein gültige Stadieneinteilung von malignen Tumoren. Dabei steht „T“ für Tumor und beschreibt die Ausdehnung des Primärtumors. „N“ bedeutet Nodus, indiziert das Auftreten regionärer Lymphknotenmetastasen. „M“ steht für Metastasen und zeigt das Auftreten von Fernmetastasen an. Durch Hinzufügen von Zahlen wird Auskunft über die anatomische Ausdehnung des malignen Prozesses gegeben. Die TNM-Einteilung der gynäkologischen Tumoren wurde so definiert, dass sie mit der FIGO-Nomenklatur übereinstimmt. Die TNM-Klassifikation wird beim Zervixkarzinom häufiger für die histologische Stadieneinteilung verwendet, wird dafür dann als pTxNxMx angegeben. Eine zusammenfassende Darstellung der Tumorstadien gemäß beider Nomenklaturen ist in Tabelle 4 wiedergegeben.

Darüber hinaus existieren histologische Klassifikationen, die auch als Grading bezeichnet werden und etwas über den Differenzierungsgrad der Krebszellen aussagen.

Tabelle 4 : Stadieneinteilung des Zervixkarzinoms (modifiziert nach Riethdorf et al. 1999b)

FIGO	TNM	Definition
O	T ₀ , TIS	Carcinoma in situ
IA1	T _{1a1} N ₀ M ₀	Stromainvasion ≤ 3 mm, horizontale Ausdehnung ≤ 7 mm
IA2	T _{1a2} N ₀ M ₀	Stromainvasion ≤ 5 mm, horizontale Ausdehnung ≤ 7 mm
IB1	T _{1b1} N ₀ M ₀	Tumor auf Zervix begrenzt, > FIGO IA2, ≤ 4 cm Durchmesser
IB2	T _{1b2} N ₀ M ₀	Tumor auf die Zervix begrenzt, > 4 cm Durchmesser
IIA	T _{2a} N ₀ M ₀	Tumor auf die Vagina übergreifend, aber nicht unteres Drittel
IIB	T _{2b} N ₀ M ₀	Infiltration der Parametrien, die Beckenwände werden nicht erreicht
IIIA	T _{3a} N ₀ M ₀	Ausdehnung bis ins untere Drittel der Vagina
IIIB	T _{3b} N ₀ M ₀	Ausdehnung bis zur Beckenwand und/oder Hydronephrose/stumme Niere
IIIC	T ₁ N ₁ M ₀	(Ausdehnung des Primärtumors wie oben), Lymphknotenmetastasen
	T ₂ N ₁ M ₀	
	T _{3a} N ₁ M ₀	
	T _{3b} N ₁ M ₀	
IVA	T ₄ N _x M ₀	Tumor infiltriert Blasen- und/oder Darmschleimhaut; fakultative Lymphknotenmetastasen
IVB	T _x N _x M ₁	(jede Art der Ausdehnung des Primärtumors); fakultative Lymphknotenmetastasen; Fernmetastasen

I C 2. 1. 5 Diagnostische Verfahren

Neben der Inspektion der Portio vaginalis und Spiegeleinstellung sind Zervixabstrich und Kolposkopie die beiden Verfahren, die sowohl zum Screening als auch zur Diagnostik des Zervixkarzinoms eingesetzt werden. Beide Verfahren haben sich im internationalen Vergleich unterschiedlich stark durchgesetzt und kommen in den meisten europäischen Ländern in der Routine zur Anwendung, während der Einsatz der Kolposkopie in den USA erst bei Auftreten eines suspekten oder positiven zytologischen Befundes als erweiterte Diagnostik erfolgt (Köchli et al. 1998). In Deutschland erfolgt im Rahmen der jährlichen Krebsfrüherkennungsuntersuchung ein Abstrichtest, fakultativ kann er durch eine Kolposkopie ergänzt werden, die die Sensitivität der Zytologie um 5-10% erhöhen kann (Köchli et al. 1998).

Zur Diagnosesicherung präklinischer Neoplasien ist ein Einsatz von Zytologie und Kolposkopie essentiell, der in Deutschland in Leitlinien empfohlen wird (Leitlinie der AG Zervixpathologie und Kolposkopie 2000). Darüber hinausgehende diagnostische Verfahren einschließlich klinischer Staginguntersuchungen beim manifesten invasiven Zervixkarzinom werden hier nicht dargelegt.

Der Zervixabstrich nach Papanicolaou (Pap Test)

Der Zervixabstrich zur frühzeitigen Erkennung von Zelldysplasien wurde Mitte der 40iger Jahre durch Papanicolaou und Traut eingeführt. Dabei wird unter Sicht und nach vorausgegangener Entfernung von Schleim und Zelldetritus mit geeigneten Entnahmeinstrumenten (z.B. spezielle Bürsten oder Spatel) zelluläres Material separat von Ekto- und Endozervix entnommen (Riethdorf et al. 1999a). Dieses Zellmaterial wird auf einen Objektträger ausgestrichen, durch unverzügliches Besprühen mit oder Eintauchen in eine 96%-ige Alkohollösung fixiert und nach der Papanicolaou-Methode im Labor gefärbt. Anschließend wird das Präparat lichtmikroskopisch beurteilt, indem der Objektträger mäanderförmig nach morphologisch veränderten Zellen durchgemustert wird (Europäische Leitlinien für die Qualitätssicherung von Screening-Programmen zur Zervixkarzinomfrüherkennung 1992).

Testgütekriterien und Fehlerquellen dieser Methode werden im Kapitel I A.2.2.3 besprochen.

Die Kolposkopie

Mit dem Kolposkop, einer 1924 von Hinselmann entwickelten 3,5 – 30fach vergrößernden Lupe, werden die Portio vaginalis und die Transformationszone auf Epithelveränderungen untersucht. Nach Entnahme des Pap Abstrichs kann die diagnostische Aussagekraft der Kolposkopie durch Essigsäureprobe (Aufbringen von 2-5%-iger Essigsäure auf die Zervixschleimhaut. Diese färbt atypisches Epithel (Präkanzerosen), aber auch normales Zylinderepithel oder physiologischerweise entstandenes metaplastisches Epithel weiß. Die biologische Grundlage der Reaktion ist nicht bekannt.) und Schiller'sche Jodprobe (Zur Durchführung des Testes wird 3%ige Jodlösung (Lugol'sche Lösung) auf die Zervixschleimhaut aufgetragen. Dadurch werden glykogenhaltige Zellen, z.B. normales nicht verhornendes Plattenepithel dunkelbraun bis schwarz gefärbt. Atypisches Epithel sowie normales Zylinderepithel und metaplastisches Epithel enthalten wenig oder kein Glykogen und färben sich daher nur wenig (gelb) oder gar nicht (jod-negativ) an.) gesteigert werden (Jourda & Girardi 2000). Am Ende einer jeden Kolposkopie steht das „Grading“: anhand von Kriterien wie Farbe, Oberflächenbeschaffenheit, Gefäßabstand, Abgrenzung und Jodanfärbung wird der Schweregrad einer Läsion eingestuft.

I C 2. 1. 6 Stadiengerechte Therapie und Nachsorge

Therapie der Zervixdysplasien

In der Behandlung der Vorstufen des Zervixkarzinoms kommen, dem Schweregrad der Läsion entsprechend, zwei operative Maßnahmen in Frage: destruierende Verfahren und Exzisionsverfahren (Leitlinie der AG Zervixpathologie und Kolposkopie 2000). Vorteil destruierender Verfahren ist ihre ambulante Durchführbarkeit und relativ geringe Komplikationsrate, entscheidender Nachteil ist die Gewebeerstörung, die keine weitere histologische Abklärung erlaubt. Zu den destruierenden Verfahren zählen (Baltzer et al. 2000, Köchli et al. 1998, Schneider et al. 2001) die Laservaporisation (Koagulation des Gewebes mit energiereichem Laserlicht), die Kryotherapie (Gewebedestruktion durch Kontaktvereisung mit Kühlmitteln wie Flüssigstickstoff, Lachgas oder Kohlendioxid) und die Elektrokauterisierung (Gewebedestruktion durch Hitzeapplikation).

Als Exzisionsverfahren wird die Konisation (Ausschneiden eines Gewebekegels aus der Portio vaginalis der Zervix uteri unter teilweiser bis subtotaler Entfernung der Schleimhaut des Zervikalkanals) durchgeführt. Die Durchführung erfolgt in Form einer Messer-, Laser- oder Schlingenkonisation sowie in kombinierten Techniken.

Exzisionsverfahren erfordern teilweise eine stationäre Behandlung, zeigen häufiger Komplikationen, erlauben jedoch eine genaue histologische Aufarbeitung z.B. zur Sicherstellung unauffälliger Resektatränder. Abhängig von angewandter Technik und Größe des entfernten Konus liegt das Komplikationsrisiko für Blutungen, Infektionen, sowie langfristige Folgen mit Subfertilität und erhöhter Neigung zu Früh- und Fehlgeburten bei etwa 2 – 5 % (Europäische Leitlinien für die Qualitätssicherung von Screening-Programmen zur Zervixkarzinomfrüherkennung 1992).

Für die Therapie der Zervixdysplasien gibt es keine international übereinstimmenden Leitlinien. Neben dem histologischen Schweregrad spielt bei der Therapieentscheidung z. B. die klinische Ausdehnung und Lage eine Rolle. Die aktuellen AWMF-Leitlinien (AWMF-Leitlinien Nr. 032/033 2000) –Empfehlungen zeigt Tabelle 5.

Tabelle 5: Empfehlungen zur operativen Therapie bei Zervixdysplasien

CIN-Klasse	Therapie
CIN 1 oder CIN 2	Bei kolposkopisch gesicherter Beschränkung auf Ektozervix Kontrolle nach drei Monaten; Bei Persistenz und ektozervikalem Sitz Biopsie, CO ₂ -Laservaporisation; Bei endozervikaler Ausdehnung Konisation.
CIN 3	Bei negativer Abrasio und kolposkopisch gesicherter Beschränkung auf Ektozervix CO ₂ -Laservaporisation bei Gewährleistung einer effektiven Gewebsdestruktion bis in die Tiefe der Drüsen; In allen Zweifelsfällen und bei endozervikaler Ausdehnung Konisation erforderlich

Quelle: AWMF-Leitlinien (AWMF Nr. 032/033 2000).

Therapie des Zervixkarzinoms

Die Behandlung des invasiven Zervixkarzinoms erfolgt stadienspezifisch. Tabelle 6 zeigt die TNM- und FIGO-Stadieneinteilung und eine hieraus abgeleitete Therapieempfehlung des Tumorzentrums München (1999). Da die vorliegende Arbeit der Krebsfrüherkennung gewidmet ist, wurde diese Empfehlung exemplarisch ausgewählt, um einen Kurzüberblick über mögliche Therapieoptionen beim invasiven Zervixkarzinom zu geben. Daneben existieren andere Therapieempfehlungen, die teilweise detaillierter auf individuelle Befunde eingehen.

Neben der Konisation kommen die in Tabelle 7 dargestellten Standard-Operationsverfahren in Frage (AWMF-Leitlinien 2000).

Die *primäre Strahlentherapie* ist in der Regel eine Kombination aus lokaler Kontakt- und einer perkutanen Hochvoltbestrahlung. Neuere Studienergebnisse zeigen eine Verbesserung der Heilungschancen, wenn die Strahlentherapie zusammen mit einer cisplatinhaltigen Chemotherapie durchgeführt wird. Bei Vorliegen von Risikofaktoren (z.B. Lymphknotenbefall, parametraner Befall) wird deshalb generell die Anwendung der kombinierten Radiochemotherapie mit Cisplatin (alternativ Carboplatin) empfohlen, es sei denn, Kontraindikationen wie Niereninsuffizienz, hohes Alter oder erheblich beeinträchtigter Allgemeinzustand oder anderes sprechen dagegen (AWMF-Leitlinien 2000).

Tabelle 6: Stadienadaptierte Therapie des Zervixkarzinoms in Deutschland

TNM	FIGO	Therapie
T1a1	IA1	Konisation, alternativ einfache Hysterektomie (vaginal / abdominal).
T1a2	IA2	Einfache Hysterektomie (vaginal / abdominal) oder Konisation. Risikofaktoren: Lymph- und Blutgefäßeinbruch, niedriger Differenzierungsgrad.
T1b1	IB1	Radikale Hysterektomie mit pelviner, ggf. paraaortaler Lymphadektomie (Wertheim-Meigs).
T1b2	IB2	
T2a	IIA	Radikale Hysterektomie mit pelviner, ggf. paraaortaler Lymphadektomie (Wertheim-Meigs).
T2b	IIB	Radikale Hysterektomie mit pelviner, ggf. paraaortaler Lymphadektomie (Wertheim-Meigs).
T3a	III A	Individuelles Vorgehen.
T3b	III B	
	III C	
T4	IV A	Vordere und / oder hintere Exenteration, Lymphadenektomie.
M1	IV B	Eventuell palliative chirurgische Maßnahmen/perkutane Strahlentherapie.

Quelle: Tumorzentrum München (1999) "Empfehlungen zur Diagnostik, Therapie und Nachsorge. Zervixkarzinom".

Tabelle 7: Standardoperationsverfahren bei invasivem Zervixkarzinom

Operation	Beschreibung
Lymphadenektomie	Im Rahmen der erweiterten Tumorchirurgie durchgeführte systematisch-radikale Lymphknotenentfernung. Beim Zervixkarzinom werden die Lymphbahnen und Lymphknoten medial und lateral der großen Beckengefäße (Arteria iliaca externa, Arteria iliaca communis und Arteria und Vena obturatoria) bis zum Beckenboden entfernt. Die pelvine Lymphknotenentfernung wird üblicherweise vor der radikalen Hysterektomie durchgeführt.
Hysterektomie	Entfernung der Gebärmutter je nach Tumorausbreitung einfach, erweitert oder radikal und auf vaginalem oder abdominalem Zugangsweg.
Radikaloperation nach Wertheim-Meigs	Nach Darstellung der Harnleiter Entnahme des Uterus unter Mitnahme des sich neben der Gebärmutter und Scheide befindlichen Binde- und Fettgewebes, sowie der Bänder zwischen Steißbein und Uterus und einer größeren Scheidenmanschette. Zusätzlich erfolgt eine Lymphadenektomie des kleinen Beckens entlang der großen Gefäße und mit Ausräumung der Fossa obturatoria.
Chirurgisches Staging	Teilweise praktizierte Vorgehensweise im Stadium III des Zervixkarzinoms zur Abschätzung der weiteren, individuellen Therapie in Abhängigkeit der Tumorausbreitung. Durch eine pelvine und paraaortale Lymphadenektomie wird die Existenz bzw. Lokalisation von Lymphknotenmetastasen festgestellt. Es erfolgt eine simultane Entfernung der befallenen Lymphknoten, der Primärtumor wird mit einer kombinierten Strahlen-Chemotherapie behandelt.
Exenteration	Ausgedehnte bis totale Entfernung der Organe des kleinen Beckens als erweiterte Radikaloperation. Je nach Tumorbefall wird eine vordere (Harnblasenentfernung) von einer hinteren (Enddarmentfernung) Exenteration unterschieden. Bei Ausdehnung auf beide Organe muss die Entfernung sowohl der Blase als auch des Enddarms im Sinne einer totalen Exenteration erfolgen.

Die *Chemotherapie* wird in der palliativen Behandlung des Zervixkarzinoms eingesetzt, kommt aber in den letzten Jahren zunehmend auch in der adjuvanten Primärtherapie zur Anwendung, z.B. um mit einer präoperativen (neoadjuvanten) Chemotherapie Operabilität bei initial ausgedehnten Befunden herzustellen. Hier sind jedoch weitere Untersuchungen notwendig, um den Wert der verschiedenen adjuvanten Chemotherapiemethoden abschließend beurteilen zu können (Köchli et al. 1998).

Tumornachsorgekonzepte

Die Nachsorgeuntersuchung von Patientinnen mit Zervixkarzinom hat folgende Ziele:

- eine frühzeitige Erkennung von Rezidiven oder Behandlungsmisserfolgen.
- die Diagnose und Therapie postoperativer und strahlentherapeutischer Nebenwirkungen.
- die Durchführung und Überwachung ambulanter adjuvanter Therapieformen.
- die psychosoziale Betreuung und Beratung inklusive Sexual- bzw. Partnerbetreuung.
- die Hilfe bei sozialer und beruflicher Wiedereingliederung.

Nachsorgeuntersuchungen werden in den ersten drei Jahren nach Primärtherapie vierteljährlich, im vierten und fünften Jahr halbjährlich und danach jährlich empfohlen (AWMF-Leitlinien 2000) (Tabelle 8).

Tabelle 8: Nachsorgeprogramm für klinisch tumorfreie Patienten in Deutschland

Untersuchungsart	Anzahl der Untersuchung pro X Monate		
	Jahr nach Primärtherapie		
	1.-2.	3.-5.	Ab 6.
Zwischenanamnese	3	6	12
Allgemeinstatus, Gynäkologische, inkl. Rektale Untersuchung, Zytologie, Kolposkopie	3	6	12
Nierensonographie	3-6	3-6	3-6
(Vaginalsonographie)	3	6	12
Sonstige bildgebende Verfahren und Tumormarker (SCC)	Bei anamnestischen oder klinischen Hinweisen		

Quelle: Tumorzentrum München (1999). "Empfehlungen zur Diagnostik, Therapie und Nachsorge. Zervixkarzinom".

I C 2. 1. 7 Prognose des invasiven Zervixkarzinoms

Wichtigster prognostischer Indikator des invasiven Zervixkarzinoms ist das klinische Stadium zum Zeitpunkt der Diagnosestellung. Allgemein werden 5-Jahres-Überlebenszeiten von 90% – 95% im Stadium I, 50% - 70% im Stadium II, 30% im Stadium III und weniger als 20% im Stadium in der Literatur genannt (Kurman 2002). Die Abbildung 6 gibt beobachtete Überlebensraten (in %) aus dem Tumorregister München wieder.

Treten bei Patientinnen mit niederen Stadien gleichzeitig Lymphknotenmetastasen auf, ist die Überlebensrate reduziert; die Überlebensrate scheint dabei mit der Anzahl betroffener Lymphknoten zu korrelieren. Ein metastatischer Befall von Lymphknoten wird bei 8% - 25% der Patientinnen im Stadium I, bei 21% - 38% im Stadium II und bei 32% - 46% der Stadium-III-Tumoren beobachtet. Die Mehrheit der Tumorrezidive tritt innerhalb von 2 Jahren nach Initialtherapie auf (Kurman 2002).

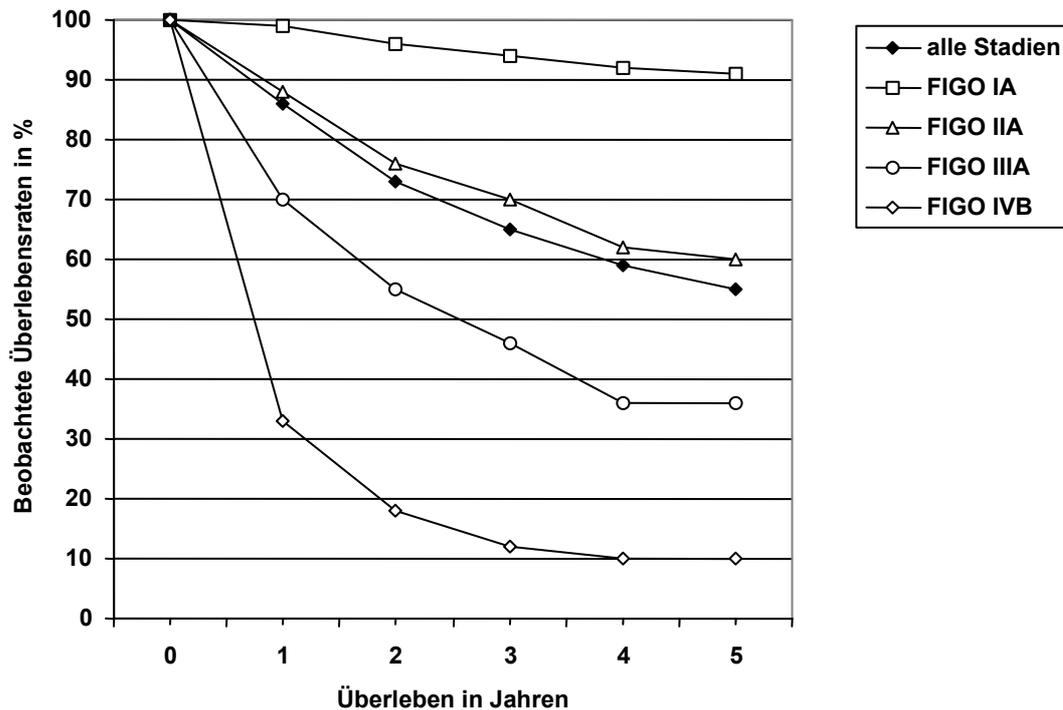


Abbildung 6: Stadienspezifische Überlebensraten beim invasiven Zervixkarzinom

Quelle: Tumorzentrum München (1996). Krebs: Häufigkeiten, Befunde und Behandlungsergebnisse: Perspektiven für die Krebsdiskussion und eine quantitative klinisch-epidemiologische Onkologie aus dem Tumorregister München. Verfasser: D. Hölzel, A. Klamert, M. Schmidt. München: Zuckschwerdt, 1996. <http://www.krebsinfo.de/ki/daten/zervix/sur1zervix.html>.

I C 2. 2 Zervixkarzinom-Screening

I C 2. 2. 1 Voraussetzungen, gesetzliche Grundlagen und Rahmenbedingungen

In Anlehnung an die Weltgesundheitsorganisation (WHO) wurden Kriterien für die Einführung von Früherkennungsprogrammen in Deutschland im §25, Absatz 3 des SGB V (2000b) festgelegt:

„Voraussetzungen für die Untersuchungen nach den Absätzen 1 und 2 (§ 25, SGB V (2000b): Gesundheitsuntersuchungen. Abs.(2): „Versicherte haben höchstens einmal jährlich Anspruch auf eine Untersuchung zur Früherkennung von Krebserkrankungen, Frauen frühestens vom Beginn des zwanzigsten Lebensjahres an,...“; Abs. (1) regelt andere Erkrankungen) ist, dass:

1. es sich um Krankheiten handelt, die wirksam behandelt werden können,
2. das Vor- oder Frühstadium dieser Krankheiten durch diagnostische Maßnahmen erfassbar ist,
3. die Krankheitszeichen medizinisch-technisch genügend eindeutig zu erfassen sind,

4. *genügend Ärzte und Einrichtungen vorhanden sind, um die aufgefundenen Verdachtsfälle eingehend zu diagnostizieren und zu behandeln.*“

Zervixkarzinome entstehen aus Präkanzerosen (Krebsvorstufen), den hochgradigen Zervixdysplasien. Der Entwicklungsprozess von Zervixdysplasien zum invasiven Zervixkarzinom ist langwierig; je nach Schweregrad der Dysplasie werden in der Literatur Zeiträume von bis zu 12 Jahren berichtet (Östor 1993 in: Broadstock 2000). Alle Stadien von Dysplasien sind kurativ behandelbar. Die „Rezidivrate“ liegt bei 3 – 6% durch unvollständig entfernte Dysplasien und persistente HPV-Infektionen (Riethdorf et al. 1999a). Bei manifesten Karzinomen ist das Stadium bei Diagnosestellung prognostisch entscheidend (Riethdorf et al. 1999b). Mit dem Pap Test können Dysplasien sowie manifeste Karzinome erkannt werden. Damit sind beim Zervixkarzinom alle Voraussetzungen für ein wirksames Screening erfüllt.

Seit 1971 sind in Deutschland Krebsfrüherkennungsuntersuchungen in den Leistungskatalog der gesetzlichen Krankenversicherung (GKV) einbezogen und als Krebsfrüherkennungsprogramm bundesweit eingeführt. Die gesetzlichen Grundlagen über die Zielpopulation und das Screeningintervall regelt §25 Abs. 2, SGB V (2000b).

Die *Richtlinien des Bundesausschusses der Ärzte und Krankenkassen über die Früherkennung von Krebserkrankungen* („Krebsfrüherkennungsrichtlinien“, zuletzt geändert 25.05.1994) bestimmen das Nähere über die den gesetzlichen Erfordernissen des §25, Abs. 2 und 3 SGB V entsprechenden ärztlichen Maßnahmen.

In der *Vereinbarung zu Qualifikationsvoraussetzungen gemäß §135 Abs. 2 SGB V zur Durchführung von zytologischen Untersuchungen zur Diagnostik der Karzinome des weiblichen Genitale* (letzte Fassung vom 01.10.1994) werden nachweispflichtige Qualifikationsvoraussetzungen von Ärzten definiert, die erforderlich sind, um zytologische Untersuchungen für die kassenärztliche Versorgung zu erbringen.

Die Ausbildung von zytologischen Assistenten ist nicht bundeseinheitlich geregelt (Schenck & Karsa 2000). Mit der *Rahmenvereinbarung über die Ausbildung und Prüfung zum technischen Assistenten/ zur technischen Assistentin an Berufsfachschulen* (Beschluss vom 12.6.1992 i.d.F. vom 26.6.1998) wurde die zuvor eigenständige Berufsausbildung zum Zytoassistenten in die Berufsausbildung der Medizin-Technischen-Assistenten (MTA) integriert. Das hatte bundesweit die Schließung staatlich betriebener Zytoassistentenschulen und eine deutliche Reduktion der Lehrinhalte gynäkologischer Zytologie zur Folge (pers. Kommunikation E. Müller-Leibinger, Vors. VDCA, 20.02.2002). Zur Sicherstellung eines Qualifikationsstandards bei zytologisch tätigen nicht-ärztlichen Mitarbeitern wurde 1998 durch die Deutsche Gesellschaft für Zytologie (DGZ) mit der freiwilligen Zertifizierung begonnen, die außer einer abgeschlossenen Berufsfachschulausbildung als Zytoassistent/MTA differenzierte Praxistätigkeiten von mindestens 1 Jahr vorsieht, um eine

Anmeldung zur Prüfung zu akzeptieren. Die Beibehaltung des Zertifikats ist an die verpflichtende Teilnahme an externen Fortbildungsveranstaltungen sowie an die Einhaltung der geltenden Richtlinien zur Qualitätssicherung gebunden (Schenck & Gassa 2001).

Rechtsverbindliche Normen der Qualitätskontrolle zytologischer Befunde, wie sie z.B. durch die CLIA (in den USA ist zur Qualitätssicherung ein 10%-iges randomisiertes manuelles Rescreening im CLIA gesetzlich vorgeschrieben. Gültige Vorschriften werden vom Clinical Laboratory Improvement Amendments, einer Behörde, die im Centers for Disease Control and Prevention (CDC) angesiedelt ist und damit eine Bundesbehörde des U.S. Departement of Health and Human Services darstellt, herausgegeben und überwacht) in den U.S.A. vorgeschrieben sind, fehlen bislang in Deutschland. In Form von Empfehlungen sind in der *Leitlinie der Bundesärztekammer zur Qualitätssicherung zytologischer Untersuchungen im Rahmen der Früherkennung des Zervixkarzinoms* (beschlossen vom Vorstand der BÄK am 17.12.1993) Standards zur Durchführung zytologischer Untersuchungen sowie zur Begutachtung und Klassifizierung des Zellmaterials (In der Leitlinie der BÄK heißt es unter 2.1.: „Die Beurteilung zytologischer Präparate obliegt dem Arzt. ...“; unter 2.2.: „Der qualifizierte Arzt kann zur Vormusterung unter seiner Aufsicht und Verantwortung ausgebildete nichtärztliche Mitarbeiter heranziehen. Die Vormusterungsleistung ...soll acht bis zehn Fälle pro Stunde nicht überschreiten. ...Mindestens bis zu 20% der negativen Befunde sind vom Arzt nachzumustern... Zytologisch auffällige, zweifelhafte und positive Präparate werden in jedem Fall vom Arzt begutachtet und klassifiziert...“) ausgegeben worden. Leider ist insbesondere zum Rescreening primär als negativ klassifizierter Befunde mit der Angabe „mindestens bis zu 20%“ keine eindeutige Empfehlung ausgesprochen worden.

I C 2. 2. 2 Das Screeningprogramm

Das seit 1971 in Deutschland etablierte *nationale Screeningprogramm* (In Deutschland werden Frauen im Rahmen des nationalen Screeningprogramms nicht aktiv zum Screening eingeladen im Gegensatz zum angelsächsischen Sprachraum, in dem bei einem systematischen Screening eine systematisch-proaktive Strategie voraussetzt wird.) (Linos & Riza 2000, Schenck & von Karsa 2000) zur Früherkennung von Gebärmutterhalskrebs wurde 1991 auf die neuen Bundesländer ausgedehnt. Seit den frühen 90-er Jahren finden Bestrebungen statt, auf dem Boden des EU-weiten Programms „Europa gegen Krebs“ mit dem dabei gegründeten Europäischen Netzwerk für Zervixkarzinom-Screening europäische Leitlinien und Qualitätssicherungsmaßnahmen für das deutsche Gesundheitssystem zu adaptieren und zu etablieren. Ziele der Maßnahmen sind die Festlegung von Standards, die sowohl die Qualität des Screenings erhöhen, als auch die Screeningabdeckung und den Res-

sourceneinsatz optimieren sollen, um Inzidenz und Mortalität des Zervixkarzinoms zu senken (Linos et al. 2000).

Für Deutschland werden für das jährliche Screening Teilnahmeraten von etwa 50 % der gesetzlich Krankenversicherten berichtet (Schenck & Karsa 2000). Für einen innereuropäischen Vergleich müssen diese Zahlen auf ein Dreijahresintervall hochgerechnet werden, da außer Deutschland nur Österreich und Luxemburg ein jährliches Screening vorsehen. Dabei liegt Deutschland hinter Finnland und Schweden, Ländern mit einem *proaktiven systematischen Screening* (auf der Basis einer meist zentralen Datenbank werden Frauen für das Screening identifiziert und angeschrieben. Die erhobenen Befunde werden zentral gespeichert und davon ausgehend werden Follow-ups festgelegt. Im angloamerikanischen Sprachraum wird für dieses Vorgehen der Begriff *Systematic Screening* verwendet) mit etwa 80% auf dem dritten Platz, aus Spanien werden Teilnahmeraten von 27% berichtet (van Ballegoien et al. 2000). Die Inanspruchnahme in Deutschland erfolgte am häufigsten in der Altersgruppe der 25- bis 54-jährigen, ab dem 55. Lebensjahr sinkt sie stetig, bei Frauen über 65 Jahren liegt sie bei 20% (Beckmann et al. 2000). Aussagen über das Teilnahmeverhalten von Frauen mit erhöhtem Zervixkarzinomrisiko, wie z.B. Frauen mit niedrigem Sozialstatus oder häufig wechselnden Geschlechtspartnern lagen uns nicht vor. Außerdem unterliegen in Deutschland die abrechnungsbezogenen Daten des Zentralinstituts der GKV über das Zervixkarzinom-Screening einem dezentralen manuellen Eingabeverfahren mit großer Variabilität der Qualität, die sich auf die Zuverlässigkeit der Daten auswirkt (Schenck & Karsa 2000).

In Deutschland werden Frauen nicht aktiv zum Screening eingeladen, eine Screeninguntersuchung erfolgt zumeist im Rahmen eines Routinebesuchs beim Gynäkologen mit informierter Zustimmung. Damit trägt das Programm de facto Züge eines *opportunistischen Screenings* (bezeichnet das Angebot eines Tests auf eine unerwartete Erkrankung zu einem Zeitpunkt, wenn eine Person einen Arzt/ eine Ärztin aus einem anderen Grunde aufsucht (Jepson et al. 2000), in diesem Falle z.B. im Rahmen der Empfängnisverhütung. Im angloamerikanischen Sprachraum werden Formen nicht-proaktiven Screenings als opportunistisch bezeichnet) (Schneider et al. 2000). Es wurde berichtet, dass opportunistische Screeningstrategien zum Over-screening bei jungen, wohlhabenden Frauen mit geringem Risiko führen und zu Underscreening älterer, weniger wohlhabender Frauen und Frauen aus Minderheitengruppen (Fahs et al. 1996 in: CCOHTA-Report 1997). In zahlreichen Studien, vorwiegend aus den U.S.A. und Kanada, sind Determinanten des Screeningverhaltens sowie Interventionen zur Erhöhung der Screeningteilnahme untersucht worden. Evidente Zusammenhänge ließen sich bislang nur für die Determinanten Alter und Teilnahmebereitschaft in vorausgegangenen Screeningtests darlegen, als eher erfolgreiche Interventionen erwiesen sich gezielte persönliche Einladungen und Arzt-Patient-Interaktionen. Für zahlreiche andere Interventionen konnte keine Evidenz

ermittelt werden, außerdem waren andere Maßnahmen selten kosteneffektiv (Jepson et al. 2000).

Etwa 90% aller Pap Tests werden in Deutschland von niedergelassenen Gynäkologen durchgeführt, die restlichen 10% von Hausärzten. Die Vergütung erfolgt gemäß dem EBM (Schenck & Karsa 2000). Die Auswertung der etwa 15 Mio. Abstriche bei GKV-Versicherten (plus 1,5 Mio. Abstriche bei PKV-Versicherten) erfolgt in etwa 2000 Labors (Ballegooijen et al. 2000) von niedergelassenen Gynäkologen, zu einem kleineren Teil bei Pathologen oder zytopathologischen Instituten.

I C 2. 2. 3 Der Ablauf der Screeninguntersuchung

Zur Verdeutlichung der Bestandteile einer Zervixkarzinom-Screeninguntersuchung soll an dieser Stelle noch einmal ihr derzeit in der Praxis üblicher Ablauf zusammengefasst dargestellt werden.

Zunächst werden im Rahmen einer gynäkologischen Untersuchung Zellen von Ekto- und Endozervix mit einem Abstrichinstrument unter Sicht entnommen und auf einem Objektträger ausgestrichen. Unmittelbar nach dem Ausstreichen muss der Abstrich fixiert werden durch Besprühen mit oder Eintauchen in eine Fixierlösung, um eine Beschädigung oder Zerstörung der Zellen zu verhindern. Der fixierte und getrocknete Objektträger wird zur Identifikation eindeutig beschriftet und mit einem Begleitbogen, der wichtige Angaben z.B. zur Anamnese enthalten sollte, zum untersuchenden Labor gesandt. Dort wird der Abstrich nach der Papanicolaou-Methode gefärbt. Der gefärbte Objektträger wird nun von einem zytologischen Assistenten mäanderförmig nach morphologisch auffälligen Zellen durchgemustert, in der Regel werden 7 Minuten Screeningzeit für einen konventionellen Pap Abstrich empfohlen. Auffällige sowie unklare Befunde werden dem zuständigen Arzt erneut zur Beurteilung vorgelegt, sowie ein gewisser Anteil unauffälliger Objektträger zum Rescreening (in der Regel etwa 10%). Die Befunde werden schriftlich niedergelegt und an die (den) einsendenden Gynäkologin(en) übermittelt, die (der) für das weitere klinische Management von Nachfolgeuntersuchungen und ggf. Therapie verantwortlich ist.

I C 2. 2. 4 Effektivität des Zervixkarzinom-Screenings

Die Wirksamkeit der organisierten Früherkennungsuntersuchungen auf Zervixkarzinom wurde niemals in randomisierten Studien nachgewiesen, jedoch haben insbesondere in den skandinavischen Ländern, wie z.B. Finnland, epidemiologische Zu-

sammenhänge zwischen der Einführung eines effektiven nationalen Screeningprogramms und einem zeitlich versetzten substantiellen Rückgang der Inzidenz und Mortalität an Zervixkarzinomen überzeugt (Boyle et al. 2000), vergleiche auch Abbildung 1 in Kapitel I C 2. 1. 2 Allerdings konnte die Prognose des Zervixkarzinoms bereits vor Einführung des Screenings verbessert werden, welches auf eine verbesserte Ausbildung und Aufklärung des medizinischen Personals, sowie eine Verbesserung der klinischen Diagnostik und Therapie zurückgeführt wurde (Schneider et al. 1999).

Kriterien, die zu einer hohen Effektivität eines Screeningprogramms beitragen, betreffen neben einer hohen Teilnahme der Zielgruppe und der Therapieoption der Zielkondition den Screeningtest selbst. Er sollte einfach, sicher und präzise durchzuführen sein und zuverlässige Ergebnisse liefern. Die Verteilung der Testergebnisse in der Zielpopulation sollte bekannt sein und es sollten geeignete cut-off Werte (Testschwellenwerte) definiert und allgemein akzeptiert sein. Der Test sollte für die Zielpopulation annehmbar sein und es sollten im Falle eines positiven Testergebnisses allgemein akzeptierte Vorgehensweisen zur weiteren Diagnostik den Betroffenen angeboten werden können (Jepson et al. 2000). Weitere Rahmenbedingungen, an denen sich die Verfahrensbewertung beim diagnostischen Test ausrichtet, werden im Kapitel I A.4.2. dargelegt.

Testgütekriterien und Zuverlässigkeit des konventionellen Pap Tests

McCrorry et al. (1999) ermittelten in einem metaanalytischen Verfahren aus insgesamt 84 eingeschlossenen Studien Schätzungen für Testgütekriterien des konventionellen Pap Tests. Dabei betragen die berechnete Gesamt-Sensitivität (Anteil der Frauen, die Veränderungen der Gebärmutter Schleimhaut haben und als solche vom Test als „positiv“ erkannt werden) 0,51 (95% CI 0,37 – 0,66) und die Gesamt-Spezifität (Anteil der Frauen, die keine Veränderungen der Gebärmutter Schleimhaut aufweisen und als solche vom Test als „negativ“ erkannt werden) 0,98 (95% CI 0,98 – 0,99). Das bedeutet, dass der Pap Test zwar in hohem Maße nur kranke Personen (Zervixdysplasien und Zervixkarzinome) als positiv anzeigt, andererseits, dass der Pap Test viele Erkrankte nicht als positiv erkennt, Erkrankungsfälle also „übersehen“ werden.

In einem weiteren Review (Nanda et al. 2000) wurden, basierend auf dem Studienpool des systematischen Reviews von McCrorry et al. (McCrorry & Matchar 1999) Sensitivitäten und Spezifitäten disaggregiert für (1) Populationen mit geringem a priori Risiko für CIN oder Zervixkarzinom, (2) verschiedene Test-Schwellenwerte und (3) für Verification Bias korrigiert dargestellt. Die Metaanalyse ergab eine Sensitivität des konventionellen Pap-Tests von 0,47 für Populationen mit geringem a priori Risiko an dem für Deutschland vergleichbaren Schwellenwert LSIL+, die noch unterhalb der

von McCrory et al. ermittelten Sensitivität von 0,51 liegt. In diese Metaanalyse wurden nur Studien eingeschlossen, die zytologische Ergebnisse mit einem histologischen Referenzstandard verifizierten. Die Ergebnisse dieses systematischen Reviews sind als valide zu betrachten und damit als auf den deutschen Kontext des Zervixkarzinom-Screenings übertragbar anzusehen. Diese ermittelten Testgütekriterien beziehen sich jedoch nur auf einen einzelnen diagnostischen Test.

In der Literatur werden mathematische Modellberechnungen zur medizinischen Effektivität eines Screeningverfahrens, die ein wiederholtes Screening einer Frau in bestimmten Zeitabständen berücksichtigen, berichtet. Durch Simulation des Krankheitsverlaufs und Zusammenführung von verschiedenen Schlüsseldaten werden in komplexen analytischen Modellen Effektivitäten für verschiedene Screeningintervalle berechnet. Sie liefern Schätzungen wie z.B. die Unterschiede in der kumulativen Zervixkarzinominzidenz, bezogen auf die Gesamtlebenszeit bei verschiedenen Screeningintervallen. Die IARC (International Agency for Research on Cancer, Organisation, die der WHO angehört) führte bereits 1986 mit Daten von acht europäischen Staaten und verschiedenen kanadischen Provinzen eine Modellrechnung durch. Dabei wurde bei jährlichem Screening eine kumulative Inzidenz von 105/100.000 ermittelt, bei dreijährigem Screeningintervall lag sie bei 157/100.000, was eine Zunahme von 50% bedeutet, in einer anderen Studie wurde ein dreifach erhöhtes Risiko, an Zervixkarzinom zu erkranken, für Frauen mit dreijährigem Screeningintervall gegenüber jährlich durchgeführten Pap Tests berichtet (Colgan et al. 2001). In zahlreichen entscheidungsanalytischen Modellrechnungen wurden nur marginale Effekte des jährlichen Screenings gegenüber einem dreijährigen Intervall gezeigt, die jedoch mit einer erheblichen Kostenerhöhung verbunden waren (Paul et al. 1991 in: Broadstock 2000). Als weiterer Gesichtspunkt werden in der noch nicht abgeschlossenen Diskussion differenzierte Screeningintervalle in Abhängigkeit vom Nachweis von HPV-Hochrisikotypen angeführt (Beckmann et al. 2000) sowie praktische Überlegungen zum Screeningverhalten und den Teilnahmeraten, die sich unter veränderten Intervallen nachteilig auf die Effektivität des Gesamtscreeningprogramms auswirken könnten.

Ursachen für falsch negative Befunde beim konventionellen Pap Test

Die Mehrzahl der Frauen, die ein invasives Zervixkarzinom entwickeln, wurden niemals zuvor oder in einem Abstand von mehr als fünf Jahren zuletzt gescreent (McCrory et al. 1999, Broadstock 2000). Falsch negative Screeningresultate, d.h. negative Testergebnisse im Pap Test trotz des Vorhandenseins von abnormen Zellen bei der untersuchten Frau, können prinzipiell aus drei möglichen Fehlerquellen resultieren, die jeweils etwa ein Drittel der Fehler ausmachen (McCrory et al. 1999):

1. Abstrichentnahmefehler, z.B. durch Wahl ungeeigneter Abstrichinstrumente;
2. Fehler bei Ausstrich und Fixierung des Abstrichs;
3. Detektionsfehler im Labor.

Abstrichentnahmefehler

Abstrichentnahmefehler haben verschiedene Ursachen. Einerseits können starke Verunreinigungen der Objektträger mit störendem Material durch einen ungünstigen Entnahmezeitpunkt, z.B. durch starke Blutbeimengungen während der Menstruation, durch eine falsche Entnahmetechnik, z.B. durch ungenügende Entfernung von Schleim und Zelldetritus, aber auch durch ungeeignete Entnahmeinstrumente entstehen.

In einer Metaanalyse von Martin-Hirsch et al. (1999) über die Effektivität verschiedener Abstrichentnahmeinstrumente konnten erhebliche Unterschiede dargestellt werden. So waren Spatelformen mit verlängerter Spitze (*Extended-tip*) dem Ayre-Spatel deutlich überlegen (Odds Ratio 2,07 bis 2,25 je nach untersuchter Population) und eine Kombination aus Spatel und Zytobürste war der Kombination aus Spatel und Watteträger sogar mit einer Odds Ratio von 3,71 – 3,88 bei der Gewinnung endozervikalen Zellmaterials überlegen und hinsichtlich der Rate adäquater Abstriche, die daraus gewonnen wurden. Auch ausgehend vom Endpunkt Detektion von Dysplasien zeigten sich *Extended-tip* Spatel und Cervexbrush™ den Ayre-Spateln überlegen (Martin-Hirsch et al. 1999). Auch in einer nicht randomisierten klinischen Studie, in der verschiedene Spatelformen, Watteträger und Abstrichschlinge miteinander verglichen wurden, schnitten Watteträger und Schlinge u.a. hinsichtlich ihrer Effektivität zur Gewinnung endozervikaler Zellen (Kohlberger et al. 1998) schlechter ab.

Beim Einsatz neuer Abstrichentnahmeinstrumente ist auch die Handhabung des Instruments, also die Erfahrung des Abstrichnehmers im Umgang damit entscheidend für die Abstrichqualität. So erfordert z.B. der Cervexbrush™, ein Instrument mit einem speziell geformten Bürstenkopf an der Spitze, dessen längere Borsten in die Endozervix hineinragen und dessen kürzere Borsten den Ektozervixbereich abstreichen, zwar nur einen Arbeitsgang für die Abstrichentnahme, er gilt jedoch als deutlich traumatischer in der Anwendung (Martin-Hirsch et al. 2000).

Ausstrichfehler und Fehler beim Fixieren des Materials

Ausstrichfehler und Fehler beim Fixieren des Materials, i.d.R. durch ein zu spätes Benetzen mit Fixierlösung, können zur Zerstörung von Zellen oder zu Artefaktbildungen führen, die Fehlinterpretationen begünstigen oder eine Auswertung unmöglich machen, was eine erneute Abstrichentnahme mit den damit verbundenen Kosten erforderlich macht.

Detektionsfehler im Labor

Das Hauptproblem für Detektionsfehler im Labor mit daraus resultierenden falsch negativen Befunden ist das Übersehen positiver Zellen. Falsch negative Abstriche haben meist wenige auffällige Zellen, die klein, unterfärbt und randständig sind (Mitchell & Medley 1995). Neben dem Übersehen pathologischer Zellen existieren Probleme bei der Interpretation.

Die Probleme bei der Interpretation von Pap Abstrichen sind vielfältig. Einerseits liegt, wie bereits im Kapitel I C 2. 1. 4 beschrieben, der Klassifikation zytologischer Befunde ein Kontinuum zugrunde, in das auch halbquantitative Merkmale wie z.B. „mäßig dysplastische Zellen“ eingehen, die auch immer vom subjektiven Eindruck des Untersuchers abhängen. Andererseits birgt eine Vielzahl von differentialdiagnostischen Zellbildern die Gefahr einer Fehlinterpretation und schließlich existieren eine Vielzahl von artifiziellen Zellveränderungen. Beispielhaft seien hier die so genannten Mikrobiopsien aufgeführt, wie sie z.B. bei traumatischer Abstrichentnahme auftreten können. Im gefärbten Präparat treten sie als *Hyperchromatic Crowded Groups* (HCG's) auf. Diese HCG's stellen normalerweise benigne Veränderungen dar, können jedoch auch jede Form von Dysplasie oder sogar Karzinomzellen repräsentieren und gelten als eine der häufigsten Ursachen von Interpretationsfehlern (DeMay 1997). Die HCG's wurden hier nur beispielhaft aufgeführt, um potentielle Ursachen für falsch interpretierte Abstriche darzustellen.

Es ist also daher naheliegend, dass die Interobserver-Variabilität für ein solches Verfahren recht hoch ist. Gupta et al. (2001) fanden in einer kleinen Studie mit 4 Untersuchern, zwei davon langjährig erfahren, dass es eine vollständige Übereinstimmung in den zytologischen Diagnosen nur in 6/50 Fällen (12%) gab, bei weiteren 26/50 Fällen (52%) wichen die Diagnosen in mindestens einem Merkmal voneinander ab, bei 18/50 Fällen (36%) sogar in mehr als einem Merkmal. In der ALTS-Studie, in der die Reproduzierbarkeit zytologischer und histologischer Befunde in einer multizentrischen Studie geprüft wurde, wurde die Interobserver-Variabilität verschiedener Untersucher bei zytologischen Befunden als mäßig beschrieben ($\kappa = 0,46$; 95% CI 0,44 – 0,48) und zeigte sich identisch zu histologischen Befunden bei kolposkopisch gewonnenen Biopsien ($\kappa = 0,46$; 95% CI 0,43 – 0,49) (Stoler et al. 2001).

Bei einer Gruppe von 13 freiwilligen Pathologen fanden Barres & Bergeron (2000) eine Übereinstimmung der Diagnosen in 65% und ein gewichtetes Kappa (Die gewichtete Kappa-Statistik misst die Divergenz (Nichtübereinstimmung, «disagreement») von Diagnosen entsprechend der Bedeutung oder der Schwere des Irrtums. Die Kappa-Statistik gibt an, wie viel besser als der Zufall zwei Untersucher übereinstimmen. Perfekte Übereinstimmung ergibt ein $k = 1$, als gute Übereinstimmung gilt $k = 0.5$, bei rein zufälliger Übereinstimmung ist $k = 0$.) von 0,66.

Dieses Ergebnis ist jedoch mit hochmotivierten Teilnehmern, die davon wussten, dass es sich um eine Studie handelt, erzielt worden, und es ist anzunehmen, dass es dadurch eher zu einer Überschätzung der Übereinstimmung im Vergleich zu Alltagsbedingungen gekommen ist.

Darüber hinaus zeigte sich in einer Studie von Lee et al. (2002) ein unterschiedlich hoher Grad an Reproduzierbarkeit der zytologischen Befunde in Abhängigkeit von dem Schweregrad der Läsion. So wurden HSIL-Befunde mit höherer Übereinstimmung der verschiedenen Untersucher diagnostiziert, am geringsten wurde die Reproduzierbarkeit von Ergebnissen bei AGUS-Befunden berichtet.

I C 2. 2. 5 Qualitätssicherung im zytologischen Labor

Maßnahmen zur Qualitätssicherung im zytologischen Labor zielen in erster Linie auf die Reduktion falsch negativer Befunde. Ein wichtiges Element stellt dafür der Prozess des *Rescreenings* dar, ein Verfahren, bei dem Objektträger mit negativem Zytologiebefund im ersten Screening einer erneuten Musterung unterzogen werden. Dafür existieren verschiedene Vorgehensweisen. In den U.S.A. wird ein 10%-iges randomisiertes Rescreening der negativen Befunde von der *CLIA* zwingend vorgeschrieben. Bei diesem Vorgehen werden allerdings schon rein rechnerisch höchstens 10% der Fälle falsch negativer Befunde entdeckt, es gilt darum inzwischen eher als laborinterne Maßnahme zur Qualitätskontrolle als eine Maßnahme zur Reduktion falsch negativer Befunde. Inzwischen wurden international verschiedene andere Verfahren entwickelt, die von einer kompletten Doppelbefundung bis zum o.gen. 10%igen Rescreening reichen, darunter auch Verfahren, wie z.B. gezieltes Rescreening bei Frauen mit Vorgeschichte positiver Befunde. Arbyn & Schenck (2000) zeigten in einer Metaanalyse aus vierzehn eingeschlossenen Studien eine deutlich verbesserte Effektivität des *Rapid Rescreening* (zeitlich begrenztes Mustern zytologischer Präparate mit einer Dauer von z.B. 30 – 90 sec. Etabliertes Verfahren in England und Schottland) gegenüber dem 10% Rescreening und Hutchinson (1996) ermittelte für dieses Verfahren die beste Kosten-Effektivität gegenüber allen anderen Methoden.

Einen anderen Ansatzpunkt bietet die Begrenzung der Arbeitsbelastung von Zytoassistenten. Während in den U.S.A. im CLIA ein gesetzlich vorgeschriebenes Limit von 12,5 Fällen pro Stunde oder 85 Fällen pro 24 Stunden (Als Fälle werden hier Patientinnen bezeichnet im Unterschied zu Objektträgern, da in manchen Praxen zwei Objektträger pro Patientin abgenommen werden.) existiert, die von einer Zytoassistentin maximal gescreent werden dürfen (Schneider; 1995), gibt es in Deutschland keine bindenden Übereinkünfte. *„Bei der in manchen Laboratorien extremen Zahl von Präparaten pro Tag und Zytologieassistentin bestehen berechtigte Zweifel, ob die zytologischen Präparate tatsächlich vollständig durchgemustert werden.“*(sic!) (Schenck; 2000). Als Antwort darauf sind elektronische Aufzeichnungssysteme wie z.B. das Pathfinder System™ (CompuCyte Corp.) oder das HOME Mikroskop basierte ARIADNE System (Schenck et al.; 2001) entwickelt worden, die nach Art eines Fahrtenschreibers den gesamten Screeningvorgang inklusive der aufgewandten Screeningzeit dokumentieren. Diese Systeme sind Dokumentationssysteme und keine Screeninggeräte im engeren Sinne und wurden daher nicht in die systematische Technologiebewertung aufgenommen.

I C 2. 2. 6 Ethische Aspekte

Gegenwärtig findet ein Wertewandel weg von einer Screeningteilnahme per se hin zu einer Teilnahme nach informierter Zustimmung statt. Das bedeutet, dass eine Frau über alle Vor- und Nachteile des Screening aufgeklärt sein muss, bevor sie einer Teilnahme am Screening zustimmen kann und dass ihre Ablehnung ebenso akzeptiert werden muss. Wahlfreiheit und Autonomie der Betroffenen können dabei durch externen Druck eingeschränkt werden (Droste 2001) und hängen wesentlich von der Qualität der Information ab. In verschiedenen Untersuchungen wurden deutliche Informationsmängel identifiziert. Laut Slater (2000) gaben 42% der von ihm an einer britischen Kolposkopieklinik befragten Frauen an, dass überhaupt kein Gespräch mit ihrem Abstrichnehmer über Vor- und Nachteile eines Pap Tests stattgefunden habe. Bei 72% der Befragten wurde nicht erwähnt, dass der Pap Test nicht 100%-ig sicher ist (Slater 2000). In einem Report über die ersten 5 Jahre des NHS Zervixkarzinom-Screeningprogramms in Großbritannien wurde berichtet, dass Frauen grundsätzlich weniger über Limitationen des Screenings unterrichtet waren, als über seine Möglichkeiten (Jepson et al. 2000). Andererseits ist beobachtet worden, dass mit zunehmender Informiertheit über mögliche Limitationen eines Tests die öffentliche Akzeptanz und die Teilnahmebereitschaft sinken (Jepson et al. 2000).

Das General Medical Council (London 1999 in Jepson et al. 2000) gab zum Informationsgehalt eines ärztlichen Aufklärungsgesprächs die Empfehlung, über Zweck des Tests, die Wahrscheinlichkeit positiver und negativer Ergebnisse, die Möglichkeit falsch positiver/negativer Ergebnisse, die Unsicherheitsfaktoren, die mit dem gesam-

ten Screeningprozess zusammenhängen, die weiteren Folgemaßnahmen inkl. ihrer Erreichbarkeit sowie sonstige medizinische, soziale oder finanzielle Auswirkungen des Screening zu informieren.

Broadstock (2000) wies insbesondere auf den ethischen Aspekt falsch positiver Befunde hin, die weitere, teilweise invasive Diagnostik nach sich ziehen und bei Betroffenen mit erheblichen Ängsten bis hin zu körperlichen Beschwerden einhergehen können. Im umgekehrten Falle, wenn die Krankheit trotz regelmäßiger Screeningteilnahme nicht verhindert werden konnte, kann bei betroffenen Frauen große Verbitterung darüber einsetzen – ein weiterer Aspekt, der dann zu befürchten ist, wenn eine möglichst hohe Screeningteilnahme per se angestrebt wird. Im First Report of the National Screening Committee (London 1998 in Jepson et al. 2000) heißt es dazu: "Wenn der Nutzen der ganzen Gemeinschaft Ziel des Screenings ist, sollte dieses Ziel nicht auf Kosten des Respekts vor einem Individuum angestrebt werden." (Übersetzung der Autoren)

I C 2. 2. 7 Ökonomische Aspekte

Ökonomische Aspekte des Zervixkarzinom-Screenings

Insgesamt wurden 1995 1,4 Mrd. DM für Früherkennungsmaßnahmen einschließlich Krebsfrüherkennungsuntersuchungen (Die hier aufgeführten Kosten beziehen alle Maßnahmen der Krebsvorsorge (z.B. Untersuchungen für Brustkrebs, Prostatakarzinome, Hautkrebs usw.) ein und sind nicht allein durch das Zervixkarzinom-Screening verursacht.) ausgegeben. Die Leistungen der GKV für die Krebsfrüherkennung allein lagen 1995 insgesamt bei 684 Mio. DM, mit einer regionalen Aufteilung von 589 Mio. DM im Westen und 95 Mio. DM im Osten (Bundesgesundheitsbericht Deutschland 1998). Dieses entspricht einem Anteil von 47,9% an den Gesamtausgaben für die gesetzliche Früherkennung. Im Vergleich zum Vorjahr wurden 9,3% mehr für Krebsfrüherkennungsuntersuchungen ausgegeben. (Bundesgesundheitsbericht Deutschland 1998). Gründe für die steigenden Kosten in der Krebsfrüherkennung werden nicht angegeben.

Ökonomische Aspekte der Zervixkarzinomerkrankung

Insgesamt wurden 1995 in Deutschland 22.977 Patientinnen mit der Diagnose Zervixkarzinom (Diese Zahl beinhaltet wahrscheinlich eine große Zahl von in-situ-Karzinomen, was die große Differenz zur Zahl gemeldeter Fälle für erbrachte statio-

näre Leistungen beim Zervixkarzinom bei den Rentenversicherern erklärt und worauf indirekt auch aus der hohen Variabilität bei der Dauer stationärer Aufenthalte geschlossen werden kann) vollstationär mit einer durchschnittlichen Verweildauer von 13 Tagen, die jedoch u.a. regional stark variierte, behandelt. Dieses entsprach 0,2% aller Krankenhausfälle in dem Jahr (Bundesgesundheitsbericht Deutschland 1998).

In den ersten drei Jahren nach einer Zervixkarzinomtherapie wird von den Trägern der gesetzlichen Kranken- und Rentenversicherung in der Regel auf Antrag im Rahmen von Rehabilitationsleistungen eine stationäre Nachsorgemaßnahme von je dreiwöchiger Dauer gewährt. Die erste Nachsorge erfolgt als Anschlussbehandlung nach Primärtherapie und die folgenden als Kuren. Das Gleiche gilt für die privaten Krankenkassen und die Beihilfestellen sowie der Sozialhilfe. Abgeschlossene stationäre medizinische Leistungen mit Erstdiagnose Zervixkarzinom wurden 1995 von den Rentenversicherungsträgern in 2.358 Fällen angezeigt. Für bösartige Neubildungen der Geschlechtorgane bei Frauen wurden 1995 im Mittel 29 Pflage tage angege ben (Bundesgesundheitsbericht Deutschland 1998).

Das Zervixkarzinom hat deutliche volkswirtschaftliche Folgen. Aufgrund Arbeitsunfähigkeit von Frauen mit der Diagnose Zervixkarzinom gingen 1994 in Deutschland 732 Erwerbstätigkeitsjahre verloren. Die Anzahl der Frührentenfälle stieg zwischen 1990 und 1995 bei den Frauen um 157%. Das Durchschnittsalter bei Berentung wegen Zervixkarzinom lag 1995 bei 46,2 Jahren, 1990 hatte das Alter noch bei 49,2 Jahren gelegen. Eine ursächliche Erklärung für diese Tatsache ist nicht bekannt, jedoch ist für diese Entwicklung nicht eine Zunahme der Erkrankungshäufigkeit verantwortlich (Bundesgesundheitsbericht Deutschland 1998).

Gesamtwirtschaftliche Ausgaben für die Zervixkarzinomtherapie wurden in der Literatur nicht gefunden.

I C 2. 3 Beschreibung der Technologie

I C 2. 3. 1 Technische Beschreibung der Testverfahren

Bis zu einem Drittel der falsch negativen Befunde bei der herkömmlichen Pap-Test Methode sind auf Interpretationsfehler durch den Zytologen zurückzuführen (Kunz 1998) (siehe auch Kapitel I C 2. 2. 3). Computergestützte Bildanalyseverfahren und computergestützte Technologien zur Unterstützung der Auswertung konventioneller Pap-Abstriche sollen die Beurteilung und Interpretation der Abstriche verbessern und

somit die Rate der falsch negativen Befunde senken. Um eine hohe Anzahl falsch negativer Befunde bei einem diagnostischen Test zu vermeiden, sollte die Sensitivität des Testes möglichst hoch sein.

Hauptproblem des herkömmlichen Verfahrens nach Papanicolaou ist, wie bereits diskutiert, die hohe Rate falsch negativer Befunde. Diese entstehen etwa zu zwei Dritteln durch falsche Entnahme- und Verarbeitungstechniken, und in etwa einem Drittel der falsch negativen Befunde werden atypische Zellen vom Zytopathologen während der Beurteilung übersehen (Hillemanns et al. 1997).

Eine Vielzahl von neuen Verfahren, die in den verschiedenen Schritten der Screeninguntersuchung eingesetzt werden können, sind entwickelt worden. Diese zielen auf eine Reduktion der falsch-negativen Rate durch qualitativ bessere, repräsentative Zellabstriche und eine optimierte (halb-) automatisierte zytopathologische Auswertung.

Tabelle 9: Einsatzbereiche von Geräten und Instrumenten im Screeningablauf

Schritt der Screeninguntersuchung	Konventionelle Methode	Neue Technologien*
Entnahme der Zellen von Ekto- und Endozervix	Holzspatel Watteträger	Cervexbrush™, verschiedene Zytobürsten und speziell geformte Kunststoffspatel
Ausstreichen auf dem Objektträger	Manuell es Ausstreichen	<u>Dünnschichtpräparationen</u> <i>AutoCyte PREP™</i>
Fixierung	Eintauchen bzw. Besprühen mit Fixierlösung	<i>ThinPrep™</i>
Färbung	Papanicolaou-Methode	Keine neue Methode von klinischer Relevanz
Untersuchung des Präparates nach Auffälligkeiten	Freies Durchmustern in mäanderförmigem Ablauf im Lichtmikroskop (Mensch)	<u>Computergestützte Bildanalyseverfahren</u> <i>AutoPap™</i> <i>AutoCyteSCREEN™</i> <i>PAPNET™</i>
		Computergestützte Dokumentation des Screeningvorgangs <i>AcCell Series 2000™</i> <i>Pathfinder System™</i>

*Kursiv gedruckt sind die Technologien, die in der vorliegenden Arbeit einer systematischen Evaluation unterzogen wurden.

Die in diesem Bericht bewerteten Technologien können, nach ihren Einsatzbereich, grundsätzlich in zwei Methoden unterteilt werden:

1. Dünnschichtpräparationen (Die Begriffe Dünnschichtpräparationen („Thin Layer“, gelegentlich auch als „Monolayer“ bezeichnet) und flüssigkeitsbasierte Präparationsstechniken („Liquid Based Cytology“) stellen Synonyme dar.)
2. Computergestützte Bildanalyseverfahren

Dünnschichtpräparationen

Dünnschichtpräparationen zielen auf eine Optimierung der Fixierung und Aufarbeitung des Zellmaterials, um hierdurch eine Verbesserung der Qualität des Zellabstriches und damit eine schnellere und zuverlässigere Untersuchung des Zellabstriches zu erlangen: schnellere und bessere Fixierung, höhere Repräsentativität des zu untersuchenden Zellmaterials, Reduktion störender Materialien wie Blut, Schleim oder Eiter. Zu der Gruppe der Dünnschichtpräparationsgeräte gehören *AutoCyte PREP™* (TriPath Imaging) und *ThinPrep™* (CYTYC Corp.). Bei beiden Methoden wird mit einem besenähnlichen Entnahmeinstrument Zellmaterial der Ekto- und Endozervix entnommen und in einer speziellen hämolytischen und mucolytischen Fixierungsflüssigkeit homogenisiert. Die weitere Verarbeitung zum Dünnschichtzellpräparat wird mit unterschiedlichen Techniken vollzogen.

AutoCyte PREP™ (TriPath Imaging) basiert auf einem Mechanismus der Dichtegradientenzentrifugation und Sedimentation. Bei *ThinPrep™* (CYTYC Corp.) dagegen werden durch einen elektrischen Dispersionsstrom Zellcluster getrennt und das Zellmaterial von Mucus getrennt. Vakuumfiltriert auf einen Polykarbonatfilter werden die Zellen schließlich in dünner Schicht auf einen Objektträger aufgebracht. *ThinPrep™*-Geräte können je nach Größe des Prozessors zwischen 62.400 und 80.000 Präparate pro Jahr verarbeiten (persönliche Kommunikation A.Ritter, Cytoc Germany GmbH i.Gr., 12.02.2002), *AutoCyte PREP™*-Geräte können etwa 50.000 Präparate pro Jahr verarbeiten (persönliche Kommunikation H. Goldbecker, Medite Medizintechnik 15.12.2001). Nach Prüfung von technischer Sicherheit und medizinischer Effektivität erteilte die FDA am 20. Mai 1996 die Marktzulassung für *ThinPrep™2000*, das *AutoCyte PREP™* System erhielt seit 12. Mai 1997 fortlaufende Premarket Approvals.

Computergestützte Bildanalyseverfahren

Zu den computergestützten Bildanalyseverfahren gehören *AutoCyte SCREEN™* (TriPath Imaging), *AutoPap™* (TriPath Imaging) und *PAPNET™* (Neuromedical Systems, Inc.), kommerziell verfügbar ist derzeit nur *AutoPap™*. Darüber hinaus ist von Cytoc Corp.® ein Gerät entwickelt worden (*ThinPrep®* Imaging System (TIS)).

Nach Herstellerangaben von Cytoc Corp.® seien klinische Studien für eine FDA-Zulassung erfolgt, mit dieser Zulassung wird für Mitte 2002 gerechnet und ein erstes Gerät soll in Deutschland im 2. Quartal aufgestellt werden (persönliche Kommunikation A. Ritter, Cytoc Germany GmbH i.Gr., 12.02.2002).

*AutoPap*TM (TriPath Imaging) ist ein computerisiertes bildverarbeitendes Gerät. Es benutzt Algorithmen, um einen Teil der Objektträger für ein manuelles Screening des Zytologen am Mikroskop auszuwählen. Hochgeschwindigkeits-Videomikroskopie, Bildinterpretations-Software und „Field-of-View“-Computer werden verwendet, um Bilder konventioneller Pap-Abstriche zu analysieren und zu klassifizieren (Halford 1998 in Broadstock 2000). Von allen, in das *AutoPap*TM-Gerät eingegebenen Objektträgern können einige Objektträger von *AutoPap*TM nicht ausgewertet werden, entweder wegen physikalischer Besonderheiten oder wegen insuffizienter Zellularität. Diese werden als *Process Review* oder *ReRun* Objektträger bezeichnet und müssen, können die Fehler nicht behoben werden, manuell ausgewertet werden. Alle übrigen (für die weiteren Zahlenangaben die Bezugsgröße 100%) werden in eine Rangfolge klassifiziert. Als „*No Further Review*“-Objektträger werden diejenigen ausgegeben, die (entsprechend Normalbefunden) keiner weiteren Überprüfung unterzogen werden müssen (bis zu 25% der verarbeiteten Objektträger). „*Review*“-Objektträger werden in einem Ranking entsprechend den Quintilen ihrer Wahrscheinlichkeit, Anomalien zu enthalten, ausgegeben, um dann von einem Zytologen unter Kenntnis dieses Rankings manuell ausgewertet zu werden (75% der verarbeiteten Objektträger). Von den Review-Objektträgern, die schließlich vom Zytologen als negativ befundet werden, im *AutoPap*TM-Ranking jedoch als am höchsten wahrscheinlich als abnormal klassifiziert wurden, werden 15% für das manuelle Qualitätskontroll-Rescreening (*QC Review*) ausgewählt. 1995 erhielt *AutoPap*TM vom FDA die vorläufige Marktzulassung als Rescreener von mindestens 10% der Pap Objektträger, die zuvor mit dem konventionellen Screeningverfahren als negativ befundet wurden. Vom FDA wurde 1998 die vorläufige Marktzulassung als Primärscreener und Qualitätskontroll-Rescreener für eine weiterentwickelte Variante von *AutoPap*TM erteilt (Broadstock 2000), am 08.10.2001 folgte die erweiterte FDA-Zulassung für die Auswertung von Dünnschichtpräparaten. Jährlich können von einem *AutoPap*TM Gerät im 24-Stunden-Betrieb (abzüglich Service- und Wartungsarbeiten und arbeitsfreier Sonntage) ca. 50.000 konventionelle Präparate oder 60.000 Dünnschichtpräparate verarbeitet werden (Angabe Fa. Medite GmbH, Generalvertrieb Deutschland für TriPath Imaging; pers. Kommunikation H. Goldbecker, 15.01.2002).

Die kommerziell nicht mehr erhältlichen, weltweit noch vereinzelt operierenden Geräte *AutoCyteSCREEN*TM und *PAPNET*TM werden nur der Vollständigkeit halber in Kurzform vorgestellt:

*AutoCyte SCREEN*TM (TriPath Imaging) akzeptierte Dünnschichtpräparate und konnte mit dem *AutoCyte PREP*TM (TriPath Imaging)-System gekoppelt werden. Mit Hilfe

eines computergestützten automatisierten Mikroskops und einer computergestützten Kamera erfolgte die Untersuchung jedes einzelnen Abstrichs. In 24 Stunden konnten 300 Abstriche pro Maschine überprüft werden (Australian Health Technology Advisory Committee (AHTAC) 1998). Das *AutoCyte SCREEN™* Gerät (TriPath Imaging) wählte nach einem statistischen Klassifikationsverfahren pro Abstrich 120 Bildausschnitte aus, die für die weitere Begutachtung auf einen Farbmonitor übertragen wurden. Der Zytologieassistent befundete dann diese 120 vorselektierten und vom Gerät für auffällig befundenen Bilder eines jeden Falles. Nach Fusion der Firmen AutoCyte und Neopath zu TriPath Imaging im Jahre 1999 wurde die Produktion von *AutoCyte SCREEN™* aus firmeninternen Gründen eingestellt, es ist kommerziell nicht mehr erhältlich (Angabe Fa. Medite GmbH, Generalvertrieb Deutschland für TriPath Imaging; pers. Kommunikation H. Goldbecker, 15.01.2002).

Bei *PAPNET™* (Neuromedical Systems Inc.) handelt es sich um ein halbautomatisches, interaktives Untersuchungsverfahren für Pap-Abstriche. An weltweit zentralisierten Scannerstandorten, wie Amsterdam, New York, Hongkong usw. wurden die eingesandten Objektträger von einem Scanner eingelesen. Pro Objektträger wurden vom Gerät bestimmte Bereiche ausgewählt und auf einem digitalen Datenträger gespeichert, der dann dezentral an einer Reader-Einheit ausgewertet wurde. Die computergestützte Auswertung erfolgte bei *PAPNET™* mit Hilfe von neuronalen Netzwerken, die in der Lage sind, aus Beispielen zu lernen und zu verallgemeinern (Mango 1994). Das System wurde nicht zur Diagnosestellung entwickelt, sondern zur Unterstützung der Zytologieassistenten bei der Erfassung von anormalen Zellen (Koss 1994). Seit dem Konkurs der Firma NSI 1999 mit Ankauf der *PAPNET™*-Patente durch TriPath Imaging ist das Gerät nicht mehr kommerziell verfügbar (Angabe Fa. Medite GmbH, Generalvertrieb Deutschland für TriPath Imaging; pers. Kommunikation H. Goldbecker, 15.01.2002). Der Standort Amsterdam, der zuletzt von einer nicht zu TriPath Imaging gehörenden Gesellschaft betrieben wurde, wurde im Jahre 2001 als aufgegeben gemeldet (Kok & Boon 2001).

Weitere technologische Entwicklungen im Zervixkarzinom-Screening

Die Bestrebungen, die Effektivität des Zervixkarzinom-Screenings zu verbessern, spiegeln sich wieder in der kontinuierlichen Entwicklung neuer Technologien. In der vorliegenden Arbeit werden diese neue Technologien nicht systematisch bewertet, jedoch sollen sie hier erwähnt werden, um einen Blick auf zukünftige Alternativen zu ermöglichen.

Einerseits werden neue Abstrichentnahmeinstrumente, wie z.B. Cervexbrush™ oder bestimmte, sogenannte Zytobürsten entwickelt, welche die Entnahmegüte des Zellmaterials verbessern sollen. Eine Reihe dieser Instrumente ist in einer Meta-

Analyse evaluiert worden, wobei ein Verbesserungspotential im Vergleich zur konventionellen Methode gezeigt werden konnte (Martin-Hirsch et al. 1999).

Einen weiteren Ansatz stellen computergestützte Systeme dar, die durch eine kontinuierliche Aufzeichnung des Untersuchungsablaufs zu einer Qualitätsverbesserung bei der konventionellen Auswertung der Papanicolaou-Abstriche beitragen sollen. Zu diesen Systemen zählt *AcCell Series 2000™* (Ampersand Medical), welches das manuelle Screening der Abstriche unterstützt. Es besteht aus einer Computereinheit, die mit den meisten, allgemein gebräuchlichen Mikroskopen kompatibel ist und es übernimmt Kontrollfunktionen ebenso wie Objektivwechsel und die elektronische Markierung auffälliger Zellen. Ein weiteres System dieser Klasse ist das *Pathfinder System™* (CompuCyte Corp.), das auch mit den gebräuchlichen Mikroskopen kombiniert werden kann. Das System zeichnet das Screeningmuster, die Größe überlappender Screeningregionen, Gebiete, die gescreent bzw. ausgelassen wurden, sowie die Screeningzeit eines jeden Objektträgers auf. Der Untersucher kann manuell elektronische Markierungen interessanter Bereiche vornehmen, die anschließend das Wiederauffinden erleichtern. Die grafische Darstellung dieses Prozesses erfolgt über einen Computermonitor. Potentieller Nutzen dieser Systeme sind die Senkung der Anzahl von Fehlinterpretationen und die Verbesserung der Reliabilität der Befundung. Diese Systeme könnten eine Rolle in der Qualitätssicherung der zytologischen Labors spielen.

Die epidemiologischen Erkenntnisse über den starken Zusammenhang zwischen einer persistierenden Infektion mit HPV und der Entstehung von Zervixkarzinomen haben eine neue Entwicklung im Zervixkarzinom-Screening hervorgerufen. Der Nachweis von Infektionen mit Hochrisiko-HPV-Typen (siehe Kapitel I C 2. 1. 3) wird z. Z. als alternative bzw. komplementäre Untersuchung zum konventionellen Papanicolaou-Test diskutiert (Schneider et al. 2001). Für den molekularbiologischen Nachweis der HPV-DNA wird auch ein Abstrich benötigt, welcher, im Gegensatz zur konventionellen Zytologie, von der Frauen selbst entnommen werden kann. Diese Tatsache könnte zu einer höheren Akzeptanz des Screenings führen und damit zu Verbesserungen der Teilnahmerate am Screeningprogramm (Dzuba et al. 2002).

Zwei weitere Geräte befinden sich derzeit noch in der klinischen Entwicklungsphase. Dabei handelt es sich um so genannte opto-elektronische Sonden, welche die Zervix direkt untersuchen, und ein anderen Ansatz als die zytologische Untersuchung darstellen. Die Polarprobe-Untersuchung (TruScan™, Polartechnics© Ltd., Sydney, Australien) misst die Antwort des Zervixepithels auf elektrische und optische Stimulation. Mit dem Antwortprofil wird das Epithel einer von drei Kategorien (Normal, Low-CIN oder High-Grade-CIN) zugeordnet. Zum gegenwärtigen Zeitpunkt sind Aussagen über die Leistungsfähigkeit dieser Innovation noch nicht möglich (CCOHTA 2001). Eine zweite Sonde, die mittels Fluoreszenz den metabolischen Status und die Struktur des Epithels identifizieren kann, wird gegenwärtig von SpectRx© and Welch-

Allyn© entwickelt, wobei sich dieses Gerät noch in einem frühen Entwicklungsstadium befindet. Hohe Erwartungen werden in die potentielle Wirksamkeit dieser Geräte wie auch in eine hohe Akzeptanz seitens der Patientinnen gesetzt (NHSC 2000).

Tabelle 10: Erwartete Vor- und Nachteile der neuen Technologien im Vergleich zum herkömmlichen Pap Test

Handelsname (Hersteller)	Funktion	Erwartete Vorteile	Erwartete Nachteile
Dünnschichtpräparationstechniken			
ThinPrep™ (CITYC Corp., Boxborough, MA)	Fixierung in hämolytisch-mucolytischer Flüssigkeit. Trennung von Zellclustern, sowie Mucus durch Dispersionsstrom. Dünnschichtapplikation über Vakuumfiltration auf Polykarbonatfilter.	Abstrich enthält weniger Verunreinigungen. Fehlinterpretationen durch Artefakte geringer. Zuverlässigere Befundung. Fixiertes Abstrichmaterial für weitere zytologische Diagnostik, z.B. HPV, nutzbar.	Informationsverlust durch Auflösung von Zellverbänden und Auswaschung von Begleitzellen (z.B. „ <i>Tumordiathese</i> “) Zusätzliche Ausbildung des Personals. Zeitintensivere Verarbeitung des Abstrichs. Zusätzliche Kosten
AutoCyte PREP™ (TriPath Imaging, Burlington, NC)	Fixierung in hämolytisch-mucolytischer Flüssigkeit. Trennung des Zellmaterials und Dünnschichtapplikation durch Dichtegradientenzentrifugation und Sedimentation	Abstrich enthält weniger Verunreinigungen. Fehlinterpretationen durch Artefakte geringer. Zuverlässigere Befundung. Fixiertes Abstrichmaterial für weitere zytologische Diagnostik, z.B. HPV, nutzbar.	Informationsverlust durch Auflösung von Zellverbänden und Auswaschung von Begleitzellen (z.B. „ <i>Tumordiathese</i> “) Zusätzliche Ausbildung des Personals. Zeitintensiver Verarbeitung des Abstrichs. Zusätzliche Kosten

Handelsname (Hersteller)	Funktion	Erwartete Vorteile	Erwartete Nachteile
Computergestützte Bildanalyseverfahren			
AutoPap™ (TriPath Imaging, Burlington, NC)	Automatisches System zur Selektion und Analyse von atypischen Zervixzellen.	Geringerer Zeitaufwand durch computergestützte Selektion. Fehlinterpretationen geringer. Zuverlässigere Befundung. Anwendung in Kombination mit AutoCyte PREP™ möglich.	Zusätzliche Ausbildung des Personals. Gerätekosten.
AutoCyte SCREEN™ (TriPath Imaging, Burlington, NC)	Früheres automatisches System zur Selektion und Analyse von atypischen Zervixzellen.	Geringerer Zeitaufwand computergestützte Selektion. Fehlinterpretationen geringer. Zuverlässigere Befundung. Anwendung in Kombination mit AutoCyte PREP™ möglich.	Zusätzliche Ausbildung des Personals. Gerätekosten
Gerät kommerziell nicht mehr verfügbar.			
PapNet™ (Neuromedical Systems, Inc., Suffern, NY)	Früheres halbautomatisches System zur Selektion und Analyse von atypischen Zervixzellen.	Nützlich bei Abstrichen, die nur wenige Zellen enthalten bzw. zur Diagnostizierung von kleinzelligen Krebsarten. Fehlinterpretationen geringer. Zuverlässigere Befundung.	Erhöhter Zeitaufwand durch Einsendung der Objektträger ins zentrale Scannerzentrum und Rücksendung von Datenträgern an die Reader-Einheiten. Zusätzliche Kosten.
Gerät kommerziell nicht mehr verfügbar.			

I C 2. 3. 2 Ökonomische Aspekte der neuen Technologien

In Deutschland werden gegenwärtig die Kosten für die jährliche Krebsvorsorgeuntersuchung mit konventionellem Pap Test von den Gesetzlichen Krankenkassen übernommen. Die Kosten der neuen Verfahren werden von den gesetzlichen Krankenkassen nicht erstattet (persönliche Auskunft Dr. B. Gibis, Kassenärztliche Bundesvereinigung, 26.10.2001). Eine Kostenerstattung durch die privaten Krankenkassen ist im Einzelfall möglich (persönliche Auskunft J. Miebach, PKV, 24.02.2003).

Zusätzliche Kosten für die Durchführung von Dünnschichtpräparationen werden in Deutschland auf ca. 10 DM (entspricht € 5,11 d.A.) geschätzt (Schenck 2000) (Beispielhaft für Dünnschichtpräparationen wird hier ein Angebot der Fa. Cytoc Corp. aufgeführt (pers. Kommunikation V. Schneider; 2002): Listenpreis für Thin-PREP™2000-Prozessor: € 50.000, bis 15.06.2002 gab es ein Spezialangebot zur Markteinführung für € 20.000. Kosten je Testkit in Abhängigkeit von der Abnahmemenge gestaffelt: von € 7,50 (bei Abnahme von 200.000 Tests) bis € 11,00 (bei Abnahme von 500 Tests.)). Bezüglich der automatisierten Bildanalyse- und unterstützenden Verfahren liegen keine Literaturangaben zur Einschätzung der zusätzlichen Kosten für Deutschland vor.

Eine Frau, die sich entschließt, einen zweiten Abstrich anfertigen zu lassen, um diesen mit automatisierten Methoden beurteilen zu lassen, erhält ein Liquidations-schreiben, um Kosten in Höhe von ca. 95 DM (entspricht € 48,57 d.A.) zu begleichen. Diese werden unter einlegendem Arzt, Zytologielabor und Firma aufgeteilt (Anonymous 1998).

In den USA werden für PAPNET™ (Neuromedical Systems Inc.) und AutoPap™ (TriPath Imaging) zwischen \$35 und \$40 zusätzlich verlangt. Die Anschaffungskosten für ein AutoPap™-Primary-System werden mit \$450.000 angegeben (American College of Obstetricians & Gynaecologists Committee on Gynaecologic Practice (ACOG) 1998).

I C 2. 4 Expertenmeinungen und Empfehlungen

Die Effektivität der Früherkennungsuntersuchung wird von der Qualität der Zellabstriche (Probensammlung und -verarbeitung), deren Interpretation und den Test-Charakteristika (Sensitivität, Spezifität) entscheidend beeinflusst. Für diese Bereiche

wurden in den letzten Jahren Technologien entwickelt, die zum einen die Qualität der Zellabstriche als auch die Interpretation der Abstriche verbessern sollen.

Da weder für die medizinische Effektivität noch für die Kosten-Effektivität der neuen Technologien in Deutschland wissenschaftliche Evidenz vorliegt, seitens der Verbraucherinnen jedoch aufgrund intensiven Industriemarketings eine vermehrte Nachfrage und Unsicherheit besteht wird eine industrie-unabhängige wissenschaftliche Evaluation der neuen Verfahren zur Beurteilung der medizinischen und gesundheitsökonomischen Wirksamkeit und Effizienz seitens der Experten zunehmend gefordert (Hepp 1998, Schenck 2000). Bis zur wissenschaftlichen Evaluation der neuen Technologien hinsichtlich ihres medizinischen Nutzens und der Kosten-Effektivität wird bisher von Seiten der Experten in Deutschland keine Empfehlung zum Einsatz der neuen Verfahren im Zervixkarzinom-Screening gegeben (Hepp 1998, Schenck 2000).

Eine Verbesserung des Zervixkarzinom-Screenings sollte jedoch durch eine Erhöhung der regelmäßigen Teilnahme an der Früherkennungsuntersuchung insbesondere auch die Beteiligung der Risikogruppen, also auch derjenigen, die noch nie oder seit längerer Zeit nicht teilgenommen haben, erreicht werden (Beckmann et al. 2000, Hillemanns et al. 1997, Küppers & Schnurch 1997, Schenck; 2000).

I C 2. 5 Lebenszyklus der Technologie

Der Lebenszyklus (Abbildung 7) einer medizinischen Technologie gliedert sich in verschiedene Entwicklungsphasen (Banta & Luce 1993, Matuszewski & Burnett 1996):

- Phase I: Unter Entwicklung, experimentelles Stadium
- Phase II: Werbephase, kurz vor der Einführung in die klinische Praxis bis Implementationsphase
- Phase III: Einsatz, Akzeptanz
- Phase IV: Niedergang
- Phase V: Technologie ist obsolet.

Bei der Beurteilung einer Technologie ist der gegenwärtige Stand im Lebenszyklus von besonderer Bedeutung. Mit zunehmendem Voranschreiten im Zyklus nehmen zum einen die zur Evaluation verfügbaren Daten, ihre Qualität und Stabilität, durch die zunehmende Erfahrung der Anwender, zu, gleichzeitig nimmt jedoch der mögliche Einfluss einer Technologiefolgenabschätzung auf die rationale Verbreitung und Anwendung der Technologie ab (Leidl 1998).

Während moderne Technologien wie Dünnschichtpräparation und automatisierte sowie halbautomatisierte Technologien im Rahmen der Zervixkarzinom-Früherkennung außerhalb Deutschlands, vor allem in den USA, schon länger zum Einsatz kommen, befinden sich diese neuen Technologien in Deutschland dagegen erst in der Einführungsphase. AutoPap™ (TriPath Imaging), PAPNET™ (Neuromedical Systems Inc.) und ThinPrep™ (CYTYC Corp.) erhielten ab 1995 offizielle Zulassungen der amerikanischen Food and Drug Administration (FDA). In anderen Ländern wurden zum Teil die Vor- und Nachteile der neuen Technologien in systematischen Reviews und Health Technology Assessments evaluiert und auch dort sind diese Technologien noch nicht im flächendeckenden Einsatz. Trotz intensiven Marketings seitens einiger Technologiehersteller wird derzeit jedoch nur in wenigen gynäkologischen Praxen Deutschlands eine Durchführung der Zervixkarzinom-Früherkennungsuntersuchung mit den neuen Technologien angeboten (persönliche Mitteilung Prof. Schenck 2001). Für die hier vorzunehmende Technologiebewertung sind folgende Geräte am Markt verfügbar: ThinPrep™ (Cytoc Corp.), AutoCyte PREP™ und AutoPap™ (beide Fa. Medite GmbH, Generalvertrieb Deutschland für TriPath Imaging). Die Kosten für diese Leistungen müssen derzeit von GKV-versicherten Frauen selbst getragen werden. Eine Akzeptanz der neuen Technologien bei Experten und Fachgesellschaften besteht zum gegenwärtigen Zeitpunkt noch nicht.

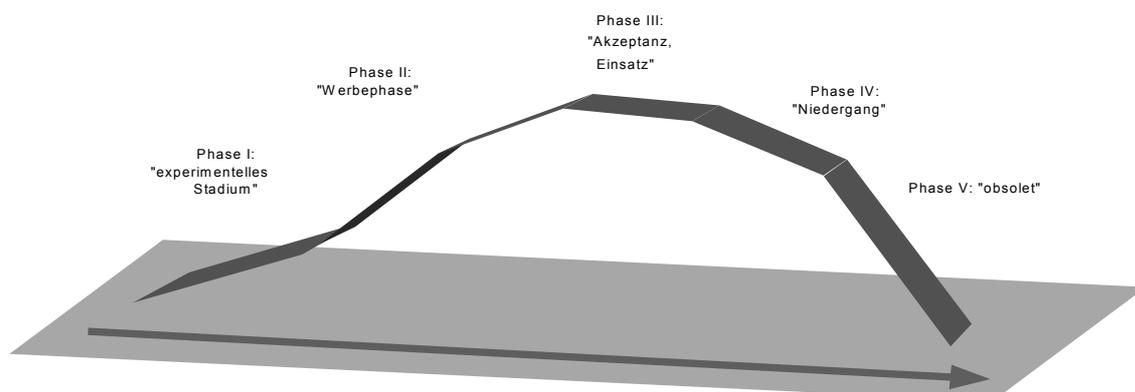


Abbildung 7: Lebenszyklus einer Technologie

I C 3 Forschungsfragen

Ziel des vorliegenden Berichtes ist eine systematische Bewertung von computergestützten Technologien auf der Basis algorithmischer Klassifikation (AutoPap™ zur Auswertung von Zervixabstrichen) sowie von Dünnschichtpräparationstechniken (AutoCyte PREP™ und ThinPrep™) zur Gewinnung flüssigkeitsbasierter Zervixabstriche im Rahmen des Krebsvorsorgeprogramms für Frauen (Eine computergestützte Auswertung auf der Basis neuronaler Netzwerke (PAPNET™) sowie das Gerät AutoCyte SCREEN™ wurden nicht eingeschlossen, da diese Geräte kommerziell nicht mehr verfügbar sind. (Siehe auch A.2.3.1.)). Die genannten Technologien kommen in Ländern außerhalb Deutschlands, vor allem in den USA, schon seit längerer Zeit zum Einsatz und erhielten von der FDA seit Mitte der neunziger Jahre ihre Marktzulassung.

Beim Zervixkarzinom-Screening geht es darum, durch Detektion von Krebsvorstufen die Inzidenz des invasiven Zervixkarzinoms (Gebärmutterhalskrebs) zu senken und durch die Erkennung von frühen Krebsstadien die Sterblichkeit an Gebärmutterhalskrebs positiv zu beeinflussen. Im weiteren soll untersucht werden, welche gesundheitsökonomischen Schlussfolgerungen sich aus einer Einführung dieser neuen Technologien in das Krebsvorsorgeprogramm für Frauen ergeben würden. Im Zentrum der Betrachtungen über die klinische Effektivität steht daher die Beantwortung folgender Fragen:

1. Wie sind anhand vorliegender Forschungsergebnisse die Genauigkeit und Zuverlässigkeit (Sensitivität, Spezifität, etc.) von Screeningtests, die mittels AutoCyte PREP™, ThinPrep™ und AutoPap™ durchgeführt werden, einzuschätzen und wie unterscheiden sie sich in diesen Testgütekriterien vom konventionellen Verfahren, dem Papanicolaou Abstrich (im folgenden Pap Test)?
2. Welche Evidenz gibt es dafür, dass der Einsatz neuer Präparationstechniken oder computergestützter Auswertungssysteme im Vergleich zu manuellem Screening von konventionellen Pap Tests durch eine höhere Entdeckungsrate von Krebsvorstufen oder frühen Krebsstadien zu einer Senkung der Inzidenz und Mortalität des invasiven Zervixkarzinoms führen kann?
3. Welche Faktoren des Zervixkarzinom-Screenings in Deutschland gewährleisten eine Übertragbarkeit internationaler Forschungsergebnisse und bei welchen Besonderheiten müssen Modifikationen in Betracht gezogen werden?
4. Welcher weitere Forschungsbedarf wird aus der Evidenzlage bei der Beantwortung der oben angeführten Forschungsfragen erkennbar?

I C 4 Methodik

Da es sich bei der vorliegenden Arbeit um eine Evaluation von Studien diagnostischer Tests handelt, bei der es im Gegensatz zur Evaluation von Therapiestudien keine allgemein anerkannte Systematik gibt und erst in jüngerer Vergangenheit eine große Zahl methodischer Instrumentarien entwickelt wurden, wird unter den jeweiligen Kapitelüberschriften auf diese besondere Problematik eingegangen.

I C 4. 1 Zielpopulation

Die Zielpopulation umfasst Frauen im Alter ab dem 20. LJ., die jedoch nicht alle in gleich hohem Maße unter dem Risiko stehen, an einem Gebärmutterhalskrebs zu erkranken (siehe Kapitel I C 2.1.3.). Jährlich erkranken in Deutschland etwa 5.800 Frauen neu an Gebärmutterhalskrebs, das entspricht einem Anteil von 3-4% aller bösartigen Neubildungen bei Frauen (Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland 1999), dabei liegt das mittlere Erkrankungsalter bei etwa 55 Jahren. Inzidenzschätzungen, die auf Daten des Saarländischen Krebsregisters beruhen, gehen von einer altersstandardisierten (bezogen auf Europabevölkerung) Erkrankungsrate über alle Altersgruppen von etwa 15 je 100.000 für das Jahr 1998 aus (Robert-Koch-Institut 1999). Diese Angaben sind im Kapitel I 2.1.2 ausführlich dargestellt. In Bezug auf die Zielpopulation wird damit ein Bereich markiert, der in der vorliegenden Arbeit als durchschnittliches Risiko aufgefasst wird.

I C 4. 2 Ergebnisparameter

I C 4. 2. 1 Systematische Konzeption in der Evaluation diagnostischer Studien

Eine komplette Evaluation eines diagnostischen Tests umfasst seine Genauigkeit, seine Zuverlässigkeit, seine Bedeutung im Gesamtkonzept diagnostischer und therapeutischer Maßnahmen in Bezug auf die Zielkondition sowie in Bezug auf das Gesamtoutcome von Patienten (Deeks 2001). Zur Beantwortung dieser Fragestellungen empfiehlt sich die Verwendung eines Stufenkonzepts, in das sowohl die Entwicklungsphasen der untersuchten Technologie als auch die spezifischen methodischen Probleme bei der Durchführung diagnostischer Studien Eingang finden. So schlugen Sackett & Haynes (2002) ein - in Anlehnung an das im Bereich therapeuti-

scher Studien übliche - 4-Phasen-Konzept vor, bestehend aus aufeinander aufbauenden Fragen, die mit den verschiedenen Studien zu beantworten sind.

Tabelle 11: Phaseneinteilung diagnostischer Studien

Phase I:	Unterscheiden sich die Testergebnisse von erkrankten Patienten gegenüber den Testergebnissen gesunder Individuen?
Phase II:	Haben Patienten mit gewissen Testergebnissen mit höherer Wahrscheinlichkeit die Zielerkrankung als Patienten mit anderen Testergebnissen?
Phase III:	Kann der Test zwischen Patienten mit und ohne Zielerkrankung unterscheiden, bei denen die Zielerkrankung klinisch vermutet wird?
Phase IV:	Haben Patienten, die getestet wurden, ein besseres Outcome als vergleichbare Patienten, die nicht getestet wurden?

Nach Sackett & Haynes, 2002

Folgt man einer derartigen Hierarchie in der Diagnostik, sind für jede der angegebenen Phasen bestimmte Studiendesigns besonders geeignet, um eine möglichst rationelle Untersuchung von diagnostischen Tests zu gewährleisten und wesentliche Fehlerquellen zu vermeiden, die zu Ergebnisverzerrungen führen.

In Tabelle 12 wird eine Gegenüberstellung von Fragestellungen und empfehlenswerten Studiendesigns gezeigt.

Tabelle 12: Studiendesigns diagnostischer Studien in Abhängigkeit zum Untersuchungsgegenstand

Untersuchungsgegenstand	Studiendesigns
Diagnostische Genauigkeit	Querschnittsstudien Fall-Kontroll-Studien Fallserien basierend auf Testergebnis Surveys an indizierter Population
Einfluss von zusätzlichen oder Ersatztests auf Prognose oder Management der Zielkondition	RCT's Kohortenstudien Fall-Kontroll-Studien "Before-and-after-"Studien

Nach Knottnerus et al.; 2002

Zusammenfassend handelt es sich bei Phase I – III Studien um Studien zur diagnostischen Genauigkeit und Zuverlässigkeit, während Phase IV Studien eine prognostische Aussage treffen. Darüber hinaus existieren andere Stufenkonzepte, so wird z.B. bei Flynn (1996) auf Stufe 1 die technische Tauglichkeit des Tests geprüft, auf Stufe 2 die diagnostische Genauigkeit, auf Stufe 3 wird geprüft, inwiefern der Test Aussa-

gen über die Wahrscheinlichkeit des Vorliegens einer Erkrankung liefert und auf Stufe 4, inwieweit der Test das weitere therapeutische Vorgehen beeinflusst. An diesen beiden Beispielen soll verdeutlicht werden, dass die Diskussion in der Literatur über das beste Vorgehen zur Evaluation diagnostischer Tests noch nicht abgeschlossen ist und dass noch kein „Goldstandard“ der Methodik dazu existiert.

Davon ausgehend wurde in der vorliegenden Arbeit eine Einschätzung der vorhandenen Studien über neue Screeningtechnologien unter den Gesichtspunkten „diagnostische Genauigkeit und Zuverlässigkeit“ und „prognostische Aussagen“ vorgenommen. Eine detaillierte Darstellung der dazu von uns geforderten Ergebnisparameter findet sich im folgenden Kapitel I A.4.2.2.

Bei der systematischen Evaluation diagnostischer Studien ist ein besonderes Augenmerk auf die Einschätzung der Studienqualität und potentieller Ergebnisverzerrungen zu richten. Lijmer et al. (1999) lieferten einige empirische Evidenz für designtypische Verzerrungen diagnostischer Studien, auf die insbesondere im Diskussionsteil näher eingegangen wird. Ergebnisverzerrungen von Studien einzuschätzen, bildet die Basis für eine Beurteilung der *internen* und *externen Validität* (Aussage darüber, in welchem Maße die Ergebnisse der individuellen Studie verzerrt sind und wie zuverlässig ihre Aussage ist).

Externe Validität: Aussage darüber, in welchem Maße die Ergebnisse individueller Studien auf andere als die Studienpopulation übertragbar sind) ihrer Ergebnisse (Deeks 2001, de Vet et al. 2001).

I C 4. 2. 2 Ergebnisparameter in der Evaluation neuer zytologischer Technologien im Zervixkarzinom-Screening

Ergebnisparameter diagnostischer Genauigkeit und Zuverlässigkeit

Grundlegende Aussagen darüber, wie gut ein Test zwischen gesunden und kranken Personen unterscheidet, liefern folgende Outcomeparameter, die in Tabelle 13 mit Definition und Berechnungsformel dargestellt werden:

Tabelle 13: Testgütekriterien

Parameter	Definition	Formel zur Berechnung innerhalb einer Vierfelder-tafel
Sensitivität*	Vermögen eines Tests, tatsächlich Erkrankte als krank zu erkennen, bzw. Anteil der erkrankten Personen, die vom Test richtig als positiv eingestuft wurden.	Sensitivität = $TP/(TP+FN)$
Spezifität*	Fähigkeit eines Tests, Gesunde richtig als gesund zu identifizieren, bzw. Anteil der nicht-erkrankten Personen, die vom Test richtig als negativ eingestuft wurden	Spezifität = $TN/(TN+FP)$
False-Positive Rate (FPR)**; [Falsch-Positiven Rate]	Anteil der falsch positiven Testergebnisse, bezogen auf alle Personen, die gesund sind.	$FPR = FP/(FP+TN)$; $FPR = 1 - (\text{Spezifität})$
False-Negative Rate (FNR)** [Falsch-Negativen Rate]	Anteil der falsch negativen Testergebnisse, bezogen auf alle Personen, die tatsächlich erkrankt sind.	$FNR = FN/(TP+FN)$; $FNR = 1 - (\text{Sensitivität})$
Positive LR (LR+)	Die LR+ vergleicht, wie viel mal häufiger bei der interessierenden Erkrankung ein richtig positiver Test als ein falsch positiver Test vorkommt. Eine „gute“ positive Likelihood Ratio ist >10 für einen positiven Test (Bestätigung der Diagnose)***	$LR+ = \text{SENS}/(1-\text{SPEC})^\dagger$
Negative LR (LR-)	Die Likelihood Ratio für einen negativen Test zum Ausschluss einer Diagnose ist kleiner 1. Eine „gute“ LR- ist <0.1.***	$LR- = (1-\text{SENS})/\text{SPEC}^\dagger$
Diagnostische Odds Ratio (DOR) [‡]	Beschreibt Chancenverhältnis eines positiven Tests bei einem Erkrankten im Verhältnis zu einem nicht Erkrankten	$DOR = (TP \times TN)/(FP \times FN)$

TP: Echt Positive, d.h. krank und positives Testergebnis;

TN: Echt Negative, d.h. nicht krank und negatives Testergebnis;

FP: Falsch Positive, d.h. nicht krank, aber positives Testergebnis

FN: Falsch Negative, d.h. krank, aber negatives Testergebnis

* nach Gordis, (2001).

** nach Svrbely J.R. & Sriram M.G.; (<http://216.157.50.80/ftp/ch39.exe>

*** Horten-Zentrum für praxisorientierte Forschung und Wissenstransfer

<http://www.evimed.ch/JournalClub/Glossar/glossar.asp?record=36>

† Medical College of Wisconsin; (<http://www.intmed.mcw.edu/clinical/bayes.html>)

‡ Nach Deeks (2001).

Die in Tabelle 13 dargestellten Testgütekriterien erfordern dichotome Testergebnisse. Liefert ein Test z.B. kontinuierliche Ergebnisse, können diese anhand festgelegter Schwellenwerte dichotomisiert werden. Die Zusammenhänge zwischen Sensitivität und Spezifität bei verschiedenen Schwellenwerten können in Form sogenannter *Receiver Operator Characteristic Curves (ROC-Kurven)* dargestellt werden. Schlechte Tests liefern Kurven, die sich nahe der Diagonalen bewegen, gute Tests weisen einen steilen Kurvenverlauf mit Annäherung an die linke obere Ecke des Diagramms auf (Deeks 2001). Zur Illustration ist ein ROC-Kurve in Abbildung 8 dargestellt. ROC-

Kurven ermöglichen auch, aus der Graphik den für den Test bestmöglichen Schwellenwert in bezug auf Sensitivität / Spezifität zu ermitteln.

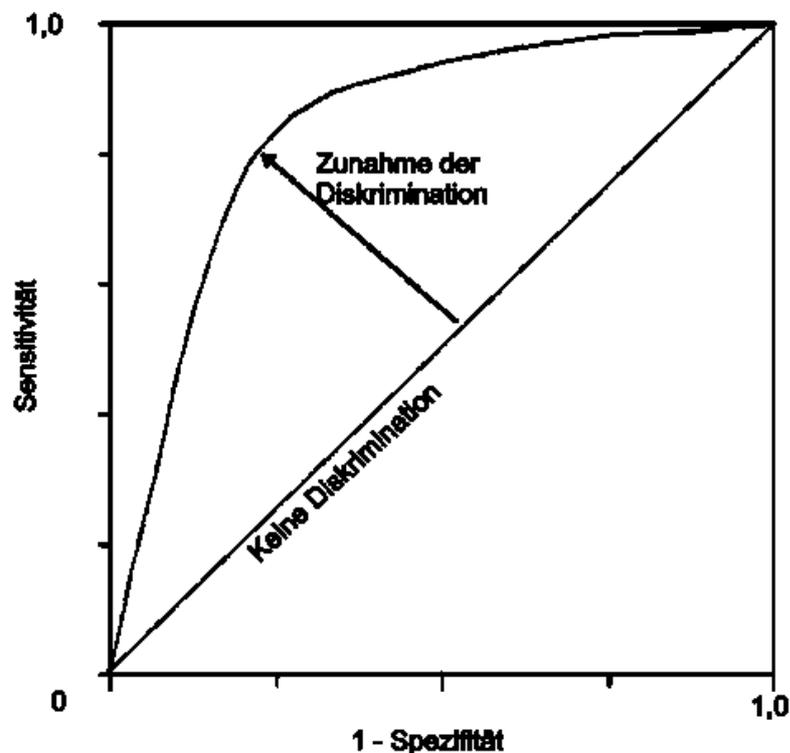


Abbildung 8: Diagramm für die Darstellung einer ROC-Kurve

Modifiziert nach Deeks, 2001.

Darüber hinaus existieren Parameter, die insbesondere im klinischen Alltag häufig benötigt werden: die Prädiktionswerte. Sie liefern Aussagen darüber, wie wahrscheinlich das Vorliegen einer Erkrankung bei positivem Testergebnis ist bzw. wie groß die Wahrscheinlichkeit ist, dass eine Person mit negativem Testergebnis gesund ist. Diese Maßzahlen sind prävalenzabhängig und schwanken darum je nach Zusammensetzung der Studienpopulation. Eine Evaluation von diagnostischen Studien unter Verwendung dieser Outcomeparameter setzt also die genaue Kenntnis der Krankheitsprävalenz voraus. Beide Maßzahlen sind in Tabelle 14 zusammengefasst.

Tabelle 14: Prädiktionswerte

Parameter	Definition	Formel zur Berechnung innerhalb einer Vierfeldertafel
Positiver Prädiktionwert (PPV)*; Synonym: positiver Vorhersagewert	Wahrscheinlichkeit dafür, dass eine Person mit einem positiven Testergebnis tatsächlich erkrankt ist. Prävalenzabhängiges Maß.	$PPV = TP/(TP+FP)$
Negativer Prädiktionwert (NPV)*; Synonym: negativer Vorhersagewert	Wahrscheinlichkeit dafür, dass eine Person mit negativem Testergebnis tatsächlich gesund ist. Prävalenzabhängiges Maß.	$NPV = TN/(TN+FN)$

TP: Echt Positive, d.h. krank und positives Testergebnis;

TN: Echt Negative, d.h. nicht krank und negatives Testergebnis;

FP: Falsch Positive, d.h. nicht krank, aber positives Testergebnis

FN: Falsch Negative, d.h. krank, aber negatives Testergebnis

* nach Gordis, (2001).

Ergebnisparameter mit prognostischer Aussage für das Patientenoutcome

Unter einem Public Health Ansatz ist über die Darstellung der Testgütekriterien hinaus von Interesse, inwiefern neue Technologien einen Beitrag zur Verbesserung des Outcome der Patienten liefern, was einem von Sackett und Haynes (2002) vorgeschlagenen Phase-IV-Ansatz entspricht. In diesem Zusammenhang wurde von uns nach Studien gesucht, denen folgende Untersuchungsgegenstände zugrunde lagen:

- Verringerung der Inzidenz von Gebärmutterhalskrebs durch die Erkennung von Krebsvorstufen;
- Verringerung der Mortalität an Gebärmutterhalskrebs durch die Erkennung früher Erkrankungsstadien.

I C 4. 3 Informationsquellen und Recherchen

In einem ersten Schritt wurden per systematische Handsuche bereits vorhandene HTA-Berichte internationaler HTA-Organisationen identifiziert und angefordert. Nach ihrer Bewertung ergab sich, dass eine Primärliteratursuche auf die Jahre 1999 bis 2001 begrenzt werden konnte. Diese Literaturrecherche wurde entsprechend einer Standard Operating Procedure (SOP) vom Deutschen Institut für Medizinische Dokumentation und Information (DIMDI) durchgeführt.

Die recherchierten Datenbanken und Suchstrategien sind im Anhang dokumentiert. Die Recherche wurde im Februar 2001 durchgeführt. Für die Publikationsjahre 1999–2001 wurde in den Datenbanken Medline, EMBASE, CancerLit und Science Citation Index recherchiert. Die identifizierten Dokumente wurden auf Dubletten geprüft und anhand von Titel und Abstract auf potentielle Relevanz für die Forschungsfragen untersucht. Die ermittelten Publikationen wurden im Volltext bestellt und einer tieferen Auswertung ihrer Relevanz und Qualität unterzogen.

Im November 2001 wurde ein Literatur-update in den Datenbanken Medline, Medline Alert und CancerLit durchgeführt.

Die elektronische Suche wurde durch eine Handsuche ergänzt, basierend auf Referenzlisten einschlägiger Monographien. Zusätzlich wurden bei verschiedenen internationalen Organisationen Leitlinien-Dokumente ermittelt und es wurde Informationsmaterial der Hersteller der Geräte, das uns durch persönliche Kontakte und über deren Internetseiten zur Verfügung gestellt wurde, zur Erarbeitung des Kapitels I A.2 über allgemeine Hintergrundinformationen berücksichtigt.

Weiterhin wurde ein Referatsband der 16. Fortbildungstagung für Klinische Zytologie, 29.11. – '03.12.2001 berücksichtigt, per Handsuche wurden Rechtsvorschriften, Rahmenvereinbarungen und allgemeine Pressemitteilungen ermittelt.

Die Hersteller Cytoc Corp®, TriPath Imaging® und die Gesellschaft für Medizintechnik mbh Medite® wurden zu Marktabdeckung, Entwicklungsprognosen aus ihrer Sicht und Gerätespezifika befragt, die über allgemeine Herstellerinformationen hinausgingen.

Als besonderes methodisches Problem bei diagnostischen Studien stellt sich eine fehlende einheitliche designbezogene Verschlagwortung dar (Deeks 2001).

I C 4. 4 Ein- und Ausschlusskriterien

In die Auswertung der klinischen Effektivität wurden HTA-Berichte, systematische Reviews und Metaanalysen über computergestützte Auswertung von Pap Abstrichen und/oder Dünnschichtpräparationstechniken einbezogen. Unsystematische Übersichten (traditionelle Reviewartikel bzw. nicht ausdrücklich als "systematisch" bezeichnete Reviewartikel) wurden nicht berücksichtigt. Darüber hinaus wurden zwei Arten von Primärstudien berücksichtigt: prognostische Studien und Studien zu Testgütekriterien. Alle diagnostischen Primärstudien mussten eine computerisierte Auswertung von

Zervixabstrichen auf der Basis algorithmischer Klassifikation (AutoPap™) und/oder Dünnschichtpräparationen (ThinPrep™ oder AutoCyte PREP™) als primären Untersuchungsgegenstand ausweisen. Die dabei vorgenommene Evaluation der Zervixzytologie sollte im Rahmen eines Screeningverfahrens erfolgt sein und der Einsatz von AutoPap™ sollte als Primärscreener, Rescreener oder Primär- und Rescreener eindeutig definiert sein.

Für die Bewertung diagnostischer Technologien wurden nach Perleth et al. (2002) und dem HTA-Report der Agency for Health Care Policy and Research (McCroory et al. 1999) zunächst folgende Einschlusskriterien definiert: Die Kriterien wurden erfüllt, wenn folgende Fragen bejahend beantwortet werden konnten:

- Wurde in der Studie deutlich beschrieben, auf welche Weise die Stichprobe gewonnen wurde (per Zufallsverfahren, durch fortlaufenden Einschluss, durch Entscheidung der Patienten, durch Zuweisung durch Ärzte, etc.)?
- Bei diagnostischen Studien zur Ermittlung der Testgenauigkeit:
 - Wurden Ergebnisparameter berichtet, die unseren Forschungsfragen entsprachen (Sensitivität, Spezifität, Positiver Vorhersagewert, Negativer Vorhersagewert, Rate falsch positiver Befunde, Rate falsch negativer Befunde, Likelihood-Ratios, diagnostische Odds Ratios, ROC-Kurven) oder ließen sie sich aus den mitgeteilten Rohdaten ermitteln?
 - Wurde zur Verifikation der zytologischen Testergebnisse ein histologischer Referenzstandard in mindestens über der Hälfte aller Fälle (>50%) eingesetzt (für positive und negative zytologische Befunde)?
- Bei Studien zu prognostischen Fragestellungen:
 - Wurde die Studie prospektiv durchgeführt oder wurde bei retrospektivem Ansatz die designtypische Verzerrung ausreichend berücksichtigt?
 - Erlaubt das Studiendesign Aussagen zum Patientenmanagement im Ergebnis erfolgter Diagnostik und/oder einen Vergleich verschiedener diagnostischer Strategien?
 - Wurden Ergebnisparameter so dargestellt, dass sich daraus mit Hilfe von Vierfeldertafeln die Effektschätzer nachprüfen ließen?

Eine Publikation wurde ausgeschlossen, wenn sie nur als Abstract vorlag, wenn es sich um Fallbeschreibungen, persönliche Korrespondenzen oder Berichte ohne klar explizierten Methodenteil handelte. Ausgeschlossen wurden auch tierexperimentelle und andere Arbeiten aus der Grundlagenforschung.

Änderungen des Screeningintervalls oder der Strategien zur Verbesserung der Erreichbarkeit von Frauen für das Screening, Geräte, die HPV-Tests durchführen, Geräte, die zur Dokumentation des Untersuchungsablaufs dienen sowie vergleichende Untersuchungen von Abstrichentnahmeinstrumenten waren nicht Gegenstand des systematischen Reviews. Allgemeine Betrachtungen zu diesen Fragen sind Bestandteil des Einführungskapitels.

I C 4. 5 Bewertung und Auswertung der Studien

Die Qualitätsbewertung der HTA-Berichte und systematischen Übersichtsarbeiten wurde mit Hilfe von Checklisten durchgeführt, die von der German Scientific Working Group Technology Assessment for Health Care entwickelt wurden. Dieses Instrument wird nicht zur Scorebildung verwendet. Die ausgefüllten Checklisten befinden sich im Anhang.

Die Bewertung der Qualität von Primärstudien folgte den unten aufgeführten Kriterien:

Bewertung der Qualität diagnostischer Studien

Die Qualität der eingeschlossenen diagnostischen Studien wurde entsprechend einem international gebräuchlichen Schema bewertet (Flynn & Adams 1996, adaptiert von Perleth et al. 2002), das eine Klassifikation in vier Qualitätsstufen vorsieht, die wie folgt definiert sind:

- A** Studien, die auf ein breites Spektrum von Patienten angewandt werden können und die keine gravierenden methodischen Fehler enthalten:
- prospektives Design
 - ≥ 35 Patienten mit und ≥ 35 Patienten ohne Krankheit

- Patienten stammen aus einer Screeningpopulation
 - Diagnose durch angemessenen Referenz- oder Goldstandard gesichert
 - vom Referenzstandard unabhängige Auswertung aller Abstriche
- B** Nur eingeschränkt generalisierbare Studien (z. B. aufgrund einer zu kleinen Anzahl von Patienten), die zwar methodische Mängel aufweisen können; diese sind jedoch beschrieben und können hinsichtlich ihrer Bedeutung auf die Schlussfolgerungen abgeschätzt werden:
- prospektives Design
 - ≥ 35 Patienten mit und ohne Krankheit
 - eingeschränktes Patientenspektrum, z.B. Universitätsklinik
- C** Studien mit einigen methodischen Mängeln, einer geringen Zahl von Teilnehmern oder mangelhafter Berichtsqualität; retrospektive Studien.
- D** Studien mit unzureichender methodischer Qualität (z.B. fehlender adäquater Referenzstandard, Patientenherkunft nicht beschrieben, Work-up-Bias wahrscheinlich) oder nicht durch Daten belegte Aussagen.

Bewertung der Qualität prognostischer Studien

Die methodische Bewertung von Studien, die prognostische Fragestellungen untersuchten, wurde in Anlehnung an Perleth et al. (2002), basierend auf einer Checkliste der Evidence-based Medicine Working Group vorgenommen (Centre for Evidence-based Medicine) vorgenommen.

1. Studienpopulation

- Repräsentativität der Studienpopulation
- Vergleichbarkeit der Krankheitsstadien

2. *Follow-up*

- ausreichende Dauer bezogen auf die Detektion relevanter Outcomes
- Vollständigkeit

3. *Outcomes*

- Objektivität bzw. ausreichend genaue Beschreibung
- verblindete, unverzerrte Erhebung

4. *Adjustierung hinsichtlich wichtiger prognostischer Faktoren, z.B.*

- Alter
- Comorbidität
- Sozioökonomischer Status

I C 4. 6 Datenextraktion und Synthese

Die Datenextraktion aus den HTA-Berichten erfolgte durch einen der Autoren. Alle Primärstudien wurden von jeweils einem der Autoren ausgewertet, die Metaanalysen von beiden Autoren unabhängig voneinander mit anschließendem Konsensusprozess. Die Synthese der Daten erfolgte narrativ, da die identifizierten Publikationen aufgrund einer erheblichen Heterogenität eine quantitative Synthese in Form einer Metaanalyse nicht zuließen. Eine Präsentation der Ergebnisse erfolgt in beschreibender Form sowie als zusammengefasste Evidenztabelle. Bei der ausführlichen Beschreibung ist für jede Publikation eine standardisierte Berichtsform gewählt worden:

- In *HTA-Berichten* bzw. *systematischen Reviews*: Dokumenttyp und Bezugsrahmen / Konkrete Fragestellung / Zielpopulation / Methode mit Recherche und Datenextraktion / Ergebnisse / Schlussfolgerungen / Abschließende Beurteilung.

- In *Primärstudien*: Fragestellung / Patienten / Methode / Ergebnisse / Schlussfolgerung der Autoren / Abschließende Bewertung.

Bei der Datenextraktion für die Evidenztabelle wurden folgende Kriterien berücksichtigt:

- Für *HTA-Berichte und systematische Reviews*: Quelle, HTA-Organisation; Fragestellung; Methodik der Informationssynthese; Ergebnisse, Schlussfolgerungen; Kritikpunkte der Autoren; Empfehlung der Autoren; Kommentar
- Für *Primärstudien*: Quelle; Fragestellung / Design; Technologie; Studienpopulation / Prävalenzangaben; Verwendeter Referenzstandard; Untersucher; Klassifikationssystem; Schwellenwert; Testgütekriterien; Weitere Ergebnisse; Bemerkungen

I C 5 Ergebnisse

Im Kapitel der Ergebnismitteilung wird nach einer Zusammenfassung der Rechercheergebnisse zunächst aus den HTA-Berichten und systematischen Übersichtsarbeiten berichtet, in denen eine Evaluation aller hier zu bewertenden Technologien vorgenommen wurde. Im weiteren Verlauf werden die Ergebnisse zum besseren Verständnis nach inhaltlichen Schwerpunkten für jede zu bewertende Technologie getrennt berichtet. Eine weitere Aufgliederung der Ergebnisdarstellung nach dem im Methodenteil vorgestellten Evaluationskonzept – gesondert nach Phasen – wurde aus Gründen der Übersichtlichkeit zurückgestellt. Auf das mehrphasige Stufenkonzept in der Evaluation diagnostischer Studien wird bei der Diskussion der Ergebnisse Bezug genommen.

I C 5. 1 Ergebnisse der Literaturrecherche

In einer Handrecherche der HTA-Institutionen wurden vier ausführliche HTA-Berichte und ein Kurz-HTA-Bericht zu neuen Screeningtechnologien identifiziert, darüber hinaus ein HTA-Bericht zu Dünnschichtpräparationstechniken.

Bei der systematischen Literaturrecherche für die Publikationsjahre 1999 – 2001 wurde in den Datenbanken Medline, EMBASE, CancerLit und Science Citation Index recherchiert. Dabei wurden im Februar 2001 n=601 Publikationen ermittelt, nach Entfernung der Dubletten wurden die verbliebenen n=465 Publikationen anhand von Titel und Abstracts auf Relevanz geprüft. Es wurden n=47 potentiell relevante Publikationen identifiziert und im Volltext bestellt. Bei dem im November 2001 durchgeführten Literatur-Update wurden weitere n=52 Publikationen ermittelt, davon wurden nach Prüfung der Abstracts weitere n=17 potentiell relevante Publikationen im Volltext bestellt. Nach Sichtung der im Volltext bestellten Publikationen wurden ein systematischer Review mit darin enthaltener Metaanalyse und elf Primärstudien identifiziert, die für die Beantwortung der Fragestellung tatsächlich relevant waren. Nach Anwendung der Ein- und Ausschlusskriterien wurden fünf Studien ausgeschlossen, so dass sechs Primärstudien und ein systematischer Review mit Metaanalyse in die Technologiebewertung zur klinischen Effektivität in die vorliegende Arbeit eingeschlossen werden konnten.

I C 5. 2 HTA-Berichte, systematischer Review und Metaanalyse

Bei der Literaturrecherche wurden vier ausführliche HTA-Berichte sowie ein Kurz-HTA-Bericht identifiziert, in denen sowohl über computergestützte Auswertung von Pap Tests als auch über Dünnschichtpräparationstechniken berichtet wurde, darüber hinaus ein HTA-Bericht und ein systematischer Review mit darin enthaltener Metaanalyse, die sich nur auf Dünnschichtpräparationstechniken bezogen. Sie sind in der Tabelle 15 mit ihrer Quelle, dem Erscheinungsjahr und den berücksichtigten Publikationszeitspannen zusammenfassend aufgelistet. Eine Zusammenfassung der Ergebnisse der HTA-Berichte und des systematischen Reviews mit Metaanalyse ist in Tabelle 16 und Tabelle 17 dargestellt.

I C 5. 2. 1 Berücksichtigte HTA-Berichte mit systematischer Bewertung aller Verfahren

Im vorliegenden Abschnitt werden die HTA-Berichte vorgestellt, die eine umfassende Bewertung aller Verfahren vornahmen, d.h. sowohl der computergestützten Auswertung von Pap Tests als auch der Dünnschichtpräparationstechniken. Da die unterschiedlichen Verfahren in diesen HTA-Berichten keiner strengen inhaltlichen Trennung unterlagen und da die dabei bewerteten Studien häufig ähnlichen methodischen Mängeln unterworfen waren, werden sie auch in der hier vorliegenden Arbeit nicht nach verfahrenstechnischen Gesichtspunkten gesondert aufgeführt. Zur besseren inhaltlichen Abgrenzung beider Verfahren werden Ergebnisse verfahrensbezogen gesondert hervorgehoben.

HTA-Bericht: Australian Health Technology Advisory Committee (AHTAC): Review of automated and semi-automated cervical screening devices. AHTAC Publications approval No. 2416. **Canberra: Commonwealth Dept. of Health and Family Services. April 1998.**

a) Dokumenttyp und Bezugsrahmen

Bei der im folgenden beschriebenen Publikation handelt es sich um einen HTA-Bericht des Australischen HTA-Komitees, bei dem eine systematische Bewertung der automatischen und halbautomatischen Geräte, die im Zervixkarzinom-Screening zur Befundermittlung eingesetzt werden, vorgenommen werden sollte.

Tabelle 15: HTA-Berichte anderer Organisationen zur Technologiebewertung – gesamt

Organisation	Titel	Publikationszeitpunkt	Literatur berücksichtigt bis	Bemerkungen
Metaanalyse von Bernstein et al.	Liquid-based cervical cytologic smear study and conventional Papanicolaou smears: A metaanalysis of prospective studies comparing cytologic diagnosis and sample adequacy.	August 2001	Januar 1990 – April 2000	Metaanalyse von Dünnschichtpräparationstechniken
NZHTA (Neuseeland)	Effectiveness and cost effectiveness of automated and semi-automated cervical screening devices: A systematic review of the literature.	Oktober 2000	01.01.1997 – 31.05.2000	Systematische Technologiebewertung aller Verfahren.
NCCHTA (Großbritannien)	Liquid-based cytology in cervical screening: a rapid and systematic review.	Juli 2000	1996 – Februar 2000	Systematische Technologiebewertung von Dünnschichtpräparationstechniken
AHCPR (U.S.A.)	Evaluation of Cervical Cytology.	Februar 1999	Bis März 1998	Systematische Technologiebewertung aller Verfahren.
AHTAC (Australien)	Review of automated and semi-automated cervical screening devices.	April 1998	1990 - 1997	Systematische Technologiebewertung aller Verfahren.
CAHTA (Catalan Agency for HTA):	Diagnóstico precoz del cáncer cervical: conocimiento actual sobre viejas y nuevas tecnologías.	Oktober 2001	Ende 2000	Ausgeschlossen, da Kurz-HTA ohne zusätzlichen Informationsgewinn
CCOHTA (Kanada)	Assessment of Techniques for Cervical Cancer Screening.	May 1997	1985 - 1997	Ausgeschlossen, da keine systematische Technologiebewertung erfolgte

Hintergrund dieser Technologiebewertung waren eine wachsende Verfügbarkeit dieser Geräte auf dem australischen Markt und ihr Einsatz bei Screeninguntersuchungen im Rahmen des Nationalen Screening Programms, ohne dass zuvor eine wissenschaftliche Bewertung ihrer Leistungsfähigkeit stattgefunden hatte.

Als Adressaten des Berichts wurden sowohl Anbieter, Entscheidungsträger als auch Konsumenten von Gesundheitsleistungen benannt.

b) Konkrete Fragestellung

Der Bericht erfasste systematisch vier Themenschwerpunkte:

1. Bewertung von Geräten für den Einsatz im Rescreening konventioneller Pap Abstriche (Hauptfokus);
2. Bewertung von Geräten im Primärscreening Einsatz;
3. Bewertung von Geräten zur verbesserten Präparationstechnik;
4. Bewertung von Geräten zum computerisierten Monitoring von manuellen Screening Prozessen

Darüber hinaus wurden der natürliche Verlauf des Zervixkarzinoms, die gegenwärtige Praxis des Zervixkarzinom-Screenings in Australien, die Leistungen und Kosten neuer Technologien, die Wahrnehmung durch Konsumenten und die Rolle der Hausärzte umfassend dargelegt. Die Effekte neuer Technologien wurden auf Laborprozesse beschränkt und es wurde u.a. folgenden Einzelfragestellungen nachgegangen:

- Schätzung der Zahl zusätzlich zu entdeckender oder vermeidbarer Zervixkarzinome für Australien durch den Einsatz neuer Geräte
- Kosten je zusätzlich entdecktem Karzinom und Kosten pro gewonnenes Lebensjahr
- Kostenvorteil automatischer und halbautomatischer Screening Geräte im Verhältnis zu anderen Methoden der Verbesserung der Sensitivität des Screening Prozesses

c) Zielpopulation

Von den Autoren wurden ausführliche epidemiologische Angaben zu Inzidenz und Mortalität des Gebärmutterhalskrebses bei australischen Frauen gemacht und sie legten Entwicklungstrends dieser Daten unter dem Einfluss des seit den 60-er Jahren eingeführten Papanicolaou-Abstrichs sowie unter dem 1991 begonnenen Nationalen Screeningprogramms dar:

Die altersstandardisierte Inzidenzrate für Zervixkarzinome betrug 1993 11,3/100.000 Frauen, was einer Anzahl von 1002 neu entdeckten Karzinomen in diesem Jahr entsprach. Während die Inzidenz der Zervixkarzinome bei über fünfzigjährigen Frauen einen signifikanten Abwärtstrend aufwies, berichteten die Autoren über einen

leichten Anstieg bei den unter fünfzigjährigen. Insgesamt machten 1995 Zervikalkarzinome 3,3% aller Krebserkrankungen australischer Frauen aus und verursachten 335 Todesfälle. Die altersstandardisierten Sterberaten waren zwischen 1984 und 1995 kontinuierlich um 2,6 % im jährlichen Durchschnitt gefallen.

Der Bericht enthielt neben diesen Angaben über die gesamte australische Bevölkerung auch detaillierte Angaben über geographische, ethnische und sozioökonomische Disproportionen, wie z.B. eine fünffach erhöhte Inzidenz und eine achtfach erhöhte Mortalität von Aboriginiefrauen gegenüber dem Bevölkerungsdurchschnitt.

d) Methode

Recherche: Die systematische Literaturrecherche zum vorliegenden Bericht wurde von CHERE (Center for Health Economics Research and Evaluation) für den Zeitraum 1990 bis 1997 in den Datenbanken Medline und CancerLit Database (National Cancer Institute) unter ausgewählten Schlagworten unternommen. Weitere Informationen wurden durch allgemeine Suche im Internet, über Hersteller und Betreiber von Geräten sowie unter Einschluss der Submissionsunterlagen für die Geräte bei der Federal Drug Administration Office (FDA, U.S.A.) bezogen. 1996 wurde von den Autoren eine öffentliche Ausschreibung des Projekts vorgenommen, eingegangene Submissionen von Gesundheitsexperten, Konsumentengruppen sowie Industrievertretern wurden ebenfalls berücksichtigt.

Datenextraktion: Eine Selektion und Bewertung der gefundenen Studien erfolgte nach Qualitätsrichtlinien des National Health and Medical Research Council (NHMRC) von 1995 und wurde von den Autoren im weiteren als mehrstufiger Prozess beschrieben, in den nur experimentelle Studien eingeschlossen wurden. Diejenigen Studien, die in *Peer Reviewed Journals* publiziert waren, wurden als diejenigen mit höchster Evidenz eingestuft. In den Prozess der Datenextraktion flossen Kriterien über den Studientyp ein, wie z.B. ob die Studie prospektiv oder retrospektiv durchgeführt wurde, über das Studiendesign, ob z.B. die Stichprobenauswahl randomisiert erfolgte, ob eine verblindete Auswertung erfolgte, ob es Routinebedingungen des Screenings abbildete. Außerdem spielten u.a. der Stichprobenumfang, die berichteten Outcomeparameter und das Vorliegen einer histologischen Verifikation eine Rolle.

e) Ergebnisse

Für die unten einzeln aufgeführten Geräte standen zur Technologiebewertung zum Zeitpunkt der Erstellung des Berichts keine Studien der Evidenzklasse I oder II (sic!)

zur Verfügung (Vom AHTAC wurden die Evidenzklassen I und II gleichbedeutend mit Metaanalysen und randomisierten klinischen Studien verwendet.).

Dünnschichtpräparationstechniken

ThinPrep™-Studien

Im Bericht wurden 13 Studien aufgeführt, davon wiesen vier Studien einen zumindest teilweisen histologischen Referenzstandard auf. Die darin berichtete Sensitivität für ThinPrep™ lag bei James & Knapp, (1992 in: AHTAC-Report, 1998), einer Studie an Hochrisikopatientinnen bei der Schwelle CIN+ bei 56,7% (76/134) (im Vergleich dazu lag die Sensitivität des konventionellen Pap Tests bei 45,5% (61/134)). Bei Sheets et al. (1995 in: AHTAC-Report, 1998) war die Sensitivität für ThinPrep™ 74%, die des konventionellen Pap Tests bei 67%. In einer dritten Studie konnte kein signifikanter Unterschied der Sensitivität festgestellt werden. Keine der aufgeführten Studien zeigte signifikante Unterschiede in Spezifität, NPV und PPV.

In weiteren, nicht histologisch verifizierten Studien wurden erhöhte Detektionsraten für insbesondere geringgradig dysplastische Befunde durch ThinPrep™ ein hoher Grad an Übereinstimmung der Diagnosen zwischen beiden Methoden sowie eine verminderte Anzahl nicht auswertbarer Präparate (unsatisfactory) berichtet.

Weiterhin berichteten die Autoren von einer signifikanten Lernkurve der auswertenden Zytologen bei den ThinPrep™-Präparaten.

AutoCyte PREP™-Studien

Es wurden vier Studien in die Auswertung aufgenommen, histologisch verifizierte Studien gab es nicht. Eine Studie (Bishop et al., 1997 in AHTAC-Report, 1998) verwendete als Referenzstandard die zytologische Konsensusdiagnose. Darin wurde bei AutoCyte PREP™ eine Sensitivität von 86,7% und eine Spezifität von 99,1% an der Schwelle LSIL+ berichtet (zum Vergleich betragen bei konventionellen Pap Abstrichen die Sensitivität 63,6% und die Spezifität 99,7% an der Schwelle LSIL+).

Andere hier berichtete Outcomeparameter bezogen sich auf den Grad an Übereinstimmung der Diagnosen zwischen beiden Methoden (84% resp. 99%) und auf die Rate nicht auswertbarer Objektträger (wie bei ThinPrep™ vermindert).

Computergestützte Auswertung von Pap Tests

AutoPap™-Studien

Von den vier im Bericht angeführten AutoPap™ Studien im Rescreeningmodus des Gerätes wurden erhöhte Detektionsraten falsch Negativer Befunde gegenüber einem 10%-igen randomisierten manuellen Rescreening beschrieben, von den Autoren wegen anderer Rescreeningvorgaben in Australien jedoch als nicht übertragbar eingeschätzt. Die Sensitivität des Rescreenings war sehr variabel, richtete sich nach der Höhe der am Gerät voreingestellten Reviewrate und reichte von 62% - 77% (10%-Reviewrate), 76% - 86% (20% Reviewrate) und 94% (30% Reviewrate) für LSIL+-Befunde in den verschiedenen Studienpopulationen.

Die einzige Studie mit histologischer Verifikation (Wilbur et al., 1996 in AHTAC-Report 1998) war schwierig zu interpretieren, da nur HSIL+-Befunde ausgewertet wurden.

f) Schlussfolgerungen

Eine Rate falsch negativer Befunde von 5 – 10% im Pap Test wird von den Autoren trotz bester Laborpraktiken als unvermeidbar angesehen. Sie kamen zu dem Schluss, dass diese Rate falsch Negativer dennoch keinen Einfluss auf den Erfolg des Screeningprogramms hat,

- weil durch ein reguläres Screening im Intervall von 2 Jahren 90% der Zervixkarzinome verhindert werden können
- weil die Prävalenz hochgradiger Dysplasien etwa 1 % beträgt und weniger als ein Drittel innerhalb der nächsten 10 Jahre ein invasives Karzinom entwickeln
- und weil nur ein geringer Teil von Todesfällen an Zervixkarzinomen einem Screeningfehler zugerechnet werden können (56% der Gestorbenen hatten niemals Pap Abstriche und nur 10% nahmen am regelmäßigen Screening teil).

Aufgrund der nur in begrenztem Ausmaß vorliegenden Evidenz über die neuen Technologien waren den Autoren keine zuverlässigen Aussagen über eine verbesserte Leistungsfähigkeit gegenüber konventionellen Pap Abstrichen möglich. Bei einem Einsatz automatischer und halbautomatischer Geräte im Zervixkarzinom-Screening kalkulierten sie einen Zuwachs zusätzlich erkannter hochgradiger Dysplasien von etwa 0,5% für die meisten australischen Laboratorien anhand der ihnen

vorliegenden Daten, auch wegen des bereits bestehenden hohen Leistungsstandards in diesen Labors.

Aus der Public Health Perspektive empfehlen sie keinen Einsatz dieser Geräte im Nationalen Screening Programm und forderten zur Durchführung bevölkerungsbezogener multizentrischer Studien auf, die Beweise für eine bessere Leistungsfähigkeit der neuen Geräte liefern müssten.

g) Abschließende Beurteilung

Im oben beschriebenen Bericht ist eine systematische Bewertung von automatischen und halbautomatischen Geräten mit Einsatzmöglichkeit im Zervixkarzinom-Screening auf dem Stand der bis 1997 recherchierten Studien vorgenommen worden. Die aufgeführten Studien wurden einer eingehenden methodischen Analyse unterzogen, ihre Ergebnisse kritisch interpretiert und in Form einer qualitativen Synthese übersichtlich präsentiert.

In den nationalen Kontext Australiens eingebunden, wurden differenzierte Empfehlungen ausgesprochen und an verschiedene Adressaten gerichtet, wurden zielgruppenspezifische Zusammenfassungen abgefasst.

Obwohl die Ergebnisse der systematischen Bewertung der berichteten Testgütekriterien unabhängig vom nationalen Kontext betrachtet werden können, sind die Schlussfolgerungen dieses Berichtes auf Deutschland nur stark eingeschränkt übertragbar,

- da sich Australien hinsichtlich der eingesetzten Screeningstrategie (das Nationale Screeningprogramm Australiens stellt eine Form des systematischen Screenings dar) von Deutschland unterscheidet, wo ein opportunistisches Screening angewendet wird.
- da das Screeningintervall in Australien zwei Jahre beträgt (in Deutschland wird jährliches Screening durchgeführt).
- da sich in Australien eine zentralisierte Infrastruktur des Screenings sowohl hinsichtlich der Laborauswertung als auch hinsichtlich der Befunddokumentation herausgebildet hat.
- da in Australien andere Richtlinien für ein Rescreening primär negativer Befunde etabliert wurden (in Australien wird überwiegend schnelles Rescreening und ge-

zieltes Rescreening eingesetzt, während in Deutschland ein 10%-iges Rescreening empfohlen wird).

- da in Australien leitliniengesteuerte Qualitätsstandards bei der Leistungsbewertung der Labors Anwendung finden, anders als in Deutschland, wo stichprobenartige Qualitätskontrollen eingesetzt werden.

HTA-Bericht: Broadstock M: Effectiveness and cost effectiveness of automated and semi-automated cervical screening devices: A systematic review of the literature. New Zealand Health Technology Assessment (NZHTA); The Clearing House for Health Outcomes and Health Technology Assessment. **NZHTA Report. 2000; 3(1).**

a) Dokumenttyp und Bezugsrahmen

Bei der vorliegenden Publikation handelt es sich um eine systematische Technologiebewertung, die als Aktualisierung an den 1998 veröffentlichten australischen HTA-Bericht (AHTAC-Report, 1998) anschließt.

Ausgehend von der Empfehlung des AHTAC-Reports, automatische und halbautomatische Verfahren im Routinescreening von Gebärmutterhalsabstrichen nicht einzusetzen - bei gleichzeitiger offensiver Werbestrategie der Gerätehersteller auf dem neuseeländischen Markt -, sollte anhand neuer Studienergebnisse die gefundene Evidenz zusammengefasst werden und eine Prüfung der Methoden hinsichtlich ihrer Eignung zur Einführung in das bevölkerungsbasierte Screeningprogramm Neuseelands unternommen werden.

Das NZHTA ist eine Forschungsabteilung der Universität von Otago, Neuseeland, finanziert vom New Zealand Health Funding Authority und dem Gesundheitsministerium.

b) Konkrete Fragestellung

Primäres Ziel des HTA-Reports war die Evaluation der klinischen Effektivität, vorzugsweise der Sensitivität und Spezifität sowie die Bewertung der Kosteneffektivität automatischer und halbautomatischer Geräte im Vergleich zu konventionellen Methoden des Zervixkarzinom-Screenings in Neuseeland.

Der Bericht sollte außerdem der Frage nachgehen, ob diese neuen Methoden für eine Einführung in das neuseeländische Screeningprogramm geeignet erscheinen.

c) Zielpopulation

Zielpopulation des bevölkerungsbezogenen Zervixkarzinom-Screenings waren nach Angaben der Autorin alle Frauen im Alter zwischen 20 und 69 Jahren. Davon ausgenommen waren Frauen, bei denen aufgrund gutartiger Veränderungen eine histologisch komplette Entfernung des Zervixepithels (z.B. im Rahmen einer Gebärmutterhalsentfernung oder einer Gebärmutterentfernung) vorausgegangen war und bei denen in Folge nur klares Sekret auftrat, d.h. bei denen es nicht z.B. zu außergewöhnlichen Blutungen kam. Die regulären Screeningintervalle wurden mit 3 Jahren angegeben, Abweichungen davon traten bei pathologischen Veränderungen, Unterbrechungen des Screeningzyklus und bei Erstuntersuchungen auf.

Das Screeningprogramm Neuseelands wurde als systematisch beschrieben, mit zentraler Registrierung der Befunde, die im weiteren für Monitoringuntersuchungen, Qualitätskontrollen und Evaluationen zur Verfügung stehen und eine Überwachung regelmäßiger Teilnahme gewährleisten, ggf. mit Aufforderung der Frauen zur Teilnahme.

d) Methode

Recherche: Über den Publikationszeitraum zwischen 01.01.1997 bis 31.05.2000 wurde eine umfangreiche systematische Literaturrecherche in zahlreichen Bibliographie- und Reviewdatenbanken, in Bibliothekskatalogen, auf Webseiten von Organisationen und als manuelle Suche in hauseigenen Dokumenten sowie nach Literaturlisten der erhaltenen Publikationen durchgeführt. Alle Schritte des Rechercheprozesses wurden detailliert beschrieben.

Eingeschlossen wurden Primärstudien über klinische Effektivitätsvergleiche neuer und konventioneller Methoden des Zervixkarzinom-Screenings, sofern über die Geräte ThinPrep™, AutoCyte PREP™ oder AutoPap™ berichtet wurde und wenn ein histologischer Referenzstandard (/oder negative Kolposkopie) oder ein zytologischer Referenzstandard, der durch ein Expertengremium festgelegt wurde, vorlag. Sekundärpublikationen wurden bei klar expliziertem Methodenteil auch dann eingeschlossen, wenn die Studienergebnisse anderen Einschlusskriterien unterworfen waren (den Einschlusskriterien der Autoren der Übersichtsarbeiten).

Ausgeschlossen wurden Publikationen in nichtenglischer Sprache, Abstracts, Fallbeschreibungen, Korrespondenzen, Studien mit unzureichender Deskription von Methoden oder Ergebnissen sowie Publikationen zu kommerziell nicht oder nicht mehr verfügbaren Geräten (z.B. PAPNET™). Es wurden Studien ausgeschlossen, wenn sie ausschließlich Untersuchungen an einer der folgenden Gruppen vornahmen: Frauen mit vormals pathologischen Abstrichen, Frauen, die noch niemals zuvor

gescreent wurden, schwangere Frauen, Frauen unter achtzehn Jahren. Weiterhin fanden Publikationen, die bereits im AHTAC-Report zitiert wurden, die über andere Geräte (HPV-testende Geräte und Location Guidance Systems (Location Guidance Systems sind Geräte, die vergleichbar einem Fahrtenschreiber den Screeningprozess aufzeichnen und damit in erster Linie zu Dokumentationszwecken aber auch zur optischen Markierung auffälliger Bereiche eingesetzt werden)) oder über Screeningstrategien berichteten und die ihre verwendeten Methoden nicht klar explizierten, keine Berücksichtigung.

Datenextraktion: Die in der Literaturrecherche identifizierten Publikationen wurden zunächst von der Autorin gesichtet und auf ihre Relevanz geprüft. Eine kritische Bewertung der eingeschlossenen Publikationen wurde nach den „Recommended Methods of the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests“ (Irwig and Glasziou, 1996) vorgenommen, die Ergebnisse wurden für jede Methode separat in Evidenztabelle zusammengefasst und einer qualitativen Synthese unterzogen.

e) Ergebnisse

Nach Sichten der Volltexte und Anwendung der Ein- und Ausschlusskriterien wurden von der Autorin 26 Publikationen ausgewählt, die in die Bewertung eingingen. Weitere 32 Publikationen wurden ausgeschlossen und in einer separaten Referenzliste aufgeführt. Gründe, die zum Ausschluss führten, wurden dabei nicht artikelbezogen angegeben.

Dünnschichtpräparationen

In die Bewertung der Dünnschichtpräparationstechniken flossen 15 primäre Forschungsarbeiten und 6 systematische Übersichtsarbeiten ein, die keine vergleichende Beurteilung der klinischen Effektivität beider Methoden (konventionelle vs. Dünnschichtmethode) nach evidenzbasierten Kriterien zuließen. Zahlen über Sensitivität und Spezifität wurden aufgrund mangelnder histologischer Referenzstandards insbesondere für negative zytologische Befunde von der Autorin nicht angegeben.

Computergestützte Auswertung von Pap Tests

Bei der Untersuchung der klinischen Effektivität des AutoPap™-Systems konnte die Autorin zwar auf 6 systematische Übersichtsarbeiten aber nur auf eine Primärstudie (Wilbur et al., 1998, 1999 in Broadstock, 2000) zurückgreifen. Da bei dieser Studie eine inkomplette Verifikation der zytologischen Befunde erfolgte, konnten die Outcomeparameter Sensitivität und Spezifität daraus nicht ermittelt werden. Erhöhte De-

tektionsraten für AutoPap™ im Rescreeningmodus im Vergleich zu konventionellem Pap Test mit zehnpotentigem randomisiertem Rescreening fanden sich nur für leichtgradige Dysplasien, nicht jedoch für hochgradige. Dieses Ergebnis wurde für Neuseeland als irrelevant erachtet, da dort andere Rescreeningpraktiken angewendet werden (gezieltes Rescreening (Wiederholte Musterung der als negativ befundenen Objektträger von Frauen mit positiven Vorbefunden oder Hochrisikoanamnese. Schnelles Rescreening (Rapid Rescreening): Schnelle Musterung aller als negativ befundenen Objektträger (ca. 30 – 90 sec. Pro Abstrich). Siehe auch Kapitel I A.2.2.4.)) und schnelles Rescreening aller Negativen).

f) Schlussfolgerungen

Die Autorin kam zu dem Schluss, dass der konventionelle Pap Test, in dreijährigem Screeningabstand bei allen Frauen eingesetzt, ausreichend wirksam ist, um 93% aller Zervixkarzinome zu verhindern, dass demgegenüber keine hinreichende Evidenz für die Überlegenheit der neuen Technologien besteht, um sie für einen Einsatz im bevölkerungsbezogenen Screening Neuseelands zu empfehlen. Sie warnte davor, Ressourcen an die Einführung neuer Geräte zu binden und schlug vor, statt dessen diese Ressourcen in eine Verbesserung der Screeningstrategien zu investieren, z.B. durch Maßnahmen zur Erhöhung der Screeningbereitschaft bei den Frauen, durch verbesserte Standards bei den Entnahmetechniken und Qualitätskontrollen sowie durch das Sicherstellen adäquater Follow-up-Untersuchungen und Therapiemaßnahmen.

g) Abschließende Beurteilung

Im HTA-Bericht der NZHTA-Organisation wurde eine umfassende und systematische Technologiebewertung der Geräte ThinPrep™, AutoCyte PREP™ und AutoPap™ für einen möglichen Einsatz im Zervixkarzinom-Screeningprogramm Neuseelands vorgenommen. Von der Autorin wurden die dazu erforderlichen methodischen Schritte klar expliziert, die Ergebnisse übersichtlich präsentiert und es erfolgte eine eingehende Diskussion methodischer Probleme der verwendeten Publikationen. Aus den Ergebnissen wurden evidenzbasierte Schlüsse gezogen und es wurde die klare Empfehlung ausgesprochen, neue Geräte nicht im Routinescreening zu implementieren, da mit den vorliegenden Forschungsergebnissen eine Überlegenheit dieser neuen Methoden nicht demonstriert werden konnte. Damit folgte die Autorin den Empfehlungen des australischen HTA- Komitees.

Die nicht in die Technologiebewertung aufgenommenen Studien wurden ohne Nennung des Ausschlussgrundes aufgelistet.

Die Übertragbarkeit auf deutsche Verhältnisse muss zurückhaltend beurteilt werden, da sich beide Länder in wesentlichen Punkten unterscheiden: hinsichtlich des Screeningprogramms (in Deutschland erfolgt de facto ein opportunistisches Screening, siehe Kapitel I A.2.2.2.), der Screeningintervalle (jährliches Screening in Deutschland) und der herrschenden Laborstandards (so erfolgt z.B. das Rescreening von negativen Befunden in Deutschland nach keiner der hier genannten Strategien, siehe Kapitel I A.2.2.4.).

HTA-Bericht: McCrory, DC, Mather DB, Bastian L, et al.: Evaluation of Cervical Cytology. Evidence Report/Technology Assessment No. 5. (Prepared by the Duke University under Contract No. 290-97-0014) **AHCPR Publication No. 99-E010. Rockville,MD: Agency for Health Care Policy and Research. February 1999.**

a) Dokumenttyp und Bezugsrahmen

Die Agency for Health Care Policy and Research (AHCPR) mit ihren Evidence-based Practice Centers (EPC's) fördert die Entwicklung von Evidenzreports und Technologiebewertungen, um öffentliche und private Organisationen bei ihrem Bemühen um Verbesserung der Qualität der medizinischen Versorgung durch Bereitstellung umfassender, wissenschaftlich fundierter Informationen zu unterstützen. Dabei werden zur Erstellung der Berichte Kooperationen zwischen den EPC's und anderen medizinischen und Forschungsorganisationen gebildet, die Berichte durchlaufen vor ihrer Publikation ein Reviewverfahren und sollen nach ihrer Veröffentlichung eine breite Gruppe von Adressaten (Leistungserbringer, Kostenträger und politische Entscheidungsträger) ansprechen.

Seit 1973 konnte ein Abfall von Inzidenz und Mortalität des Zervixkarzinoms in den U.S.A. gezeigt werden. Er war weitgehend dem Massenscreening mit dem Pap Test zur Diagnostik präkanzeröser oder früher Erkrankungsstadien zugeschrieben worden. Dennoch herrschte noch immer Ungewissheit über die Leistungsfähigkeit des Pap Tests, sowie über die Effektivität neu eingeführter Technologien, durch Reduktion der Abstrichentnahmefehler und der Screeningfehler im Labor die Rate falsch negativer Befunde zu verringern.

b) Konkrete Fragestellung

Primäres Ziel des HTA-Reports war einerseits die Determinierung der Testgüte von konventionellen Pap Tests, andererseits ihr Vergleich mit neuen Technologien (Dünnschichtpräparationstechnik, computerisiertes Rescreening auf der Basis neuronaler Netzwerke und algorithmusbasierte Klassifikationstechnologien).

Die Autoren untersuchten die diagnostische Genauigkeit der o. g. Verfahren hinsichtlich ihrer Fähigkeit, Gebärmutterhalskrebs und seine Vorstufen zu erkennen.

In zwei weiteren Fragestellungen wird gesundheitsökonomischen Aspekten nachgegangen. Dieser Teil wird im ökonomischen Teil der vorliegenden Arbeit dargestellt. (Vgl. Kapitel I A.9.)

c) Zielpopulation

Als primäre Zielpopulation wurden Frauen im Alter vom Beginn ihrer sexuellen Aktivität bis zum 85. LJ, die unter einem durchschnittlichen Risiko (sic!) stehen, an einem Zervixkarzinom zu erkranken, benannt. Die Autoren verwiesen darauf, dass ein großer Teil der Zervixkarzinome bei Frauen auftritt, die niemals oder nur begrenzt an Screeningmaßnahmen teilgenommen haben und dass weiterführende Betrachtungen zur Verbesserung der Screeningstrategie nicht Gegenstand des Berichts sind. Ebenso wurde nicht explizit auf Probleme beim Screening von älteren Frauen eingegangen. Die Autoren verwiesen dazu auf andere Publikationen.

d) Methode

Recherche: Eine systematische Literaturrecherche in den Datenbanken Medline, CancerLit, HealthSTAR, CINAHL, EMBASE und EconLit wurde im Januar 1998 durchgeführt, die jeweils bis zu den ersten Einträgen seit Bestehen dieser Datenbanken zurückverfolgt wurde, Updates erfolgten im Februar und März d.J. Darüber hinaus wurde eine Sekundärsuche nach neuen, noch nicht indexierten Artikeln und nach Herstellerinformationen auf deren Webseiten vorgenommen. Alle Schritte des Rechercheprozesses sind detailliert beschrieben.

Ein- und Ausschlusskriterien wurden für den klinischen und den gesundheitsökonomischen Part des Berichts gesondert in mehrstufigen Prozeduren entwickelt. Dabei wurde das aufwendig gestaltete Ein- und Ausschlussverfahren, das von mehreren Personen unabhängig voneinander mit anschließendem Konsensusprozess durchgeführt wurde, ausführlich dargelegt.

Es wurden diagnostische Evaluationsstudien, Metaanalysen und Reviews in den Report aufgenommen. Ein- und Ausschlusskriterien für Studien, die in die Metaanalysen und Reviews aufgenommen wurden, wurden nicht gesondert expliziert. Es lagen keine abgeschlossenen randomisierten kontrollierten Studien (RCT's) zum Vergleich konventioneller mit neuen Technologien vor.

Ausgeschlossen wurden im ersten Schritt Studien, die Zervikalzytologie nicht als Screeningtest evaluiert haben, die nicht als Primärscreening, Rescreening oder Primär- und Rescreening angelegt waren oder deren Referenzstandard nicht Histologie, Kolposkopie oder Zytologie war. Im zweiten Schritt wurden Studien ausgeschlossen, die nur einen zytologischen Referenzstandard verwendeten, bei denen der Abstand zwischen Screening Test und Referenzstandard größer als 3 Monate war und bei denen nicht alle Zellen einer Vierfeldertafel belegt waren. Da durch diese Kriterien zu wenige Studien über neue Technologien eingeschlossen werden konnten, wurde für Dünnschichtpräparationsverfahren und computergestützte Auswertungsverfahren eine Modifikation der Einschlusskriterien vorgenommen, sie wurden bei erneuter Sichtung ausgeschlossen, wenn sie kein zweiarmig prospektives Design aufwiesen, wenn keine endgültige Beurteilung diskordanter Testergebnisse durch ein Expertengremium stattfand, wenn keine histologische/kolposkopische Verifikation von wenigstens 50% der hochgradigen Dysplasien (HSIL+) vorgenommen wurde oder wenn keine Berechnung von Sensitivität (bzw. relativer Rate echt positiver Befunde) und Spezifität (bzw. relativer Rate falsch positiver Befunde) nach vorliegendem Studiendesign möglich war (Nach Vorschlag der Intersociety Working Group for Cytology Technologies (1998) und FDA (1997) ist Zytologie als Surrogatvariable eines Referenzstandards akzeptabel, wenn sie von einem unabhängigen Expertengremium im Konsensusprozess erhoben wird und wenn außerdem eine histologische Verifikation eines signifikanten Anteils der hochgradigen Dysplasien (sic!) erfolgt ist).

Datenextraktion: Die Extraktion der Daten erfolgte nach paarweisem Abgleich der Ergebnisse in den Reviewerteams als Eintrag in Evidenztabellen des AHCPR-Formats. Kreuztabellen wurden ebenfalls von den Teammitgliedern unabhängig erstellt und anschließend abgeglichen, ebenso erfolgte die Berechnung von Sensitivität, Spezifität (kalkuliert an verschiedenen Schwellenwerten), Prävalenz u.a. für die Metaanalyse benötigter Daten. Zur Bewertung der Testgüte von neuen Technologien, bei denen aus den Daten keine Berechnung von Sensitivität und Spezifität möglich war, wurden relative Raten echt Positiver/falsch Positiver (Nach Vorschlägen von Chock et al. (1997) sind Berechnungen von relativen Raten echt/falsch Positiver dann zulässig, wenn bei fehlender histologischer Verifikation von negativen zytologischen Befunden konkordant negative Befunde vorliegen und die Tests unabhängig appliziert wurden (es liegt z.B. Abhängigkeit vor, wenn initial negativ befundete Tests einem Rescreening unterzogen werden)) hinzugezogen.

Für die Qualitätsbewertung der untersuchten Studien wurde in einem Konsensusprozess der Reviewer ein Katalog von Anforderungen aufgestellt, der nachfolgend gewichtet und als Qualitätsscore für jede Studie ermittelt wurde.

Metaanalyse konventioneller Pap Tests: Für die Metaanalyse des konventionellen Pap Tests wurde die „Methode der standardisierten Differenz der empirischen Mittelwerte“ nach Hasselblad & Hedges, (1995) und Fahey et al. (1995) (beide in McCrory

et al., 1999) eingesetzt. Die dabei ermittelten Test *Effectiveness Scores* sind statistische Größen, die etwas über die Diskriminationsfähigkeit eines Tests mit nicht-dichotomen Ergebnissen aussagen (Raum & Perleth, 2002). Aus den Effectiveness Scores können Sensitivität und Spezifität abgeleitet werden. Weiterhin können aus den Effectiveness Scores *ROC-Kurven* (Receiver Operator Characteristic Curve, eine grafische Darstellung, mittels derer die Fähigkeit eines Tests ausgedrückt wird, zwischen gesunden und kranken Personen zu unterscheiden (Last 2001). Vgl. auch A.4.2.2) erstellt werden (Raum & Perleth, 2002).

Da es sich bei zytologischen Befunden um kategoriale Variablen (nicht-dichotome) handelt, bei denen zahlreiche Schwellenwerte zur Festlegung eines pathologischen Pap Testergebnisses denkbar sind, wurde hier von den Autoren ein übliches Vorgehen gewählt. Da die Studien unterschiedliche Referenzstandards einsetzten und sich auch bezüglich der untersuchten Studienpopulation deutlich voneinander unterscheiden, nahmen die Autoren Heterogenität zwischen den Studien an und verwendeten das dafür angebrachte Random-Effects-Model (Methode der Schließenden Statistik, die eine Gewichtung für eine Komponente erlaubt, um die sich Studien voneinander unterscheiden. Das Modell erlaubt auch metaanalytische Berechnungen für heterogene Studien. (Das einfachere Fixed-Effect-Model setzt Homogenität voraus.)) (Raum & Perleth, 2002).

e) Ergebnisse

Zur Bewertung der neuen Technologien wurden 6 AutoPap™-Studien, 11 PAPNET™-Studien und 8 ThinPrep™-Studien herangezogen. Da in der vorliegenden Arbeit das PAPNET™-Verfahren wegen fehlender kommerzieller Verfügbarkeit von uns ausgeschlossen wurde, wird die Bewertung der PAPNET™-Studien hier nicht expliziert.

Computergestützte Auswertung von Pap Tests

Da das Erkrankungsspektrum bei der Auswertung von Objektträgern im Rescreeningmodus (es werden Objektträger, die initial im manuellen Screening als negativ befundet wurden, verarbeitet) grundsätzlich vom Primärscreeningmodus (es werden alle Objektträger verarbeitet) verschieden ist und darum nicht miteinander vergleichbar ist, die AutoPap™ Geräte in beiden Modi mit geringfügig anderen Algorithmen arbeiten, wurden beide Modi gesondert bewertet. Aus sechs AutoPap™-Studien konnte die Sensitivität des Rescreeningmodus (Der Rescreeningmodus wird am Gerät voreingestellt. Dabei wird die Rate mit 10%, 20%, 30% etc. voreingestellt und das Gerät arbeitet mit etwas anderen Algorithmen.) eingeschätzt werden. Die Ergebnisse waren beim 20% Rescreeningmodus mit Sensitivitäten zwischen 43% und 66% an der Schwelle ASCUS+ und zwischen 66% und 92% an der Schwelle LSIL+ sehr

heterogen. Die Spezifität konnte jedoch nicht ermittelt werden. Für den Primärscreeningmodus von AutoPap™ wurden zwei Studien, die mit verschiedenen AutoPap™ Gerätetypen durchgeführt wurden, herangezogen. Darin wurde jeweils eine Sensitivität von 86% genannt, in einer der beiden Studien konnte die Spezifität auf 98% geschätzt werden.

Im weiteren wurde über die Outcomeparameter FN und FNR berichtet. In den zitierten Studien konnten mit dem AutoPap™ System sowohl im Primärscreeningmodus als auch im Rescreeningmodus im Vergleich zu konventionellen Pap Tests plus 10%iges randomisiertes Rescreening mehr falsch Negative detektiert werden, die FNR gesenkt werden. Die Beurteilung dieser Ergebnisse ist jedoch schwierig, da die Studien verschiedene methodische Mängel aufwiesen.

Dünnschichtpräparationstechniken

Von den 8 eingeschlossenen ThinPrep™-Studien konnten nur aus zwei Studien Schätzungen von Sensitivität und Spezifität abgeleitet werden. In der Studie von Bolick & Hellman (1998) wurden die Diagnosen von ThinPrep™-Objektträgern an der Schwelle LSIL+ (Schwellenwert, bei dem alle unter LSIL liegenden Befunde als negativ zusammengefasst werden und alle Befunde, die LSIL oder höhergradigen Dysplasien entsprechen, als positiv zusammengefasst werden. Siehe auch Kapitel A.2.1.4., Nomenklatur) mit dem histologischen Referenzstandard an der Schwelle CIN 2 – 3 oder darüber verglichen und erlaubten die direkte Ermittlung einer Sensitivität von 94% und einer Spezifität von 58%. In der gleichen Studie hatten konventionelle Pap Tests eine Sensitivität von 85% und eine Spezifität von 36% bei den gleichen Schwellenwerten. In der Studie von Robert et al. (1997 in: McCrory et al., 1999) wurde ein von den Autoren akzeptierter Referenzstandard nur für die positiven Befunde angewendet (incl. einer histologischen Verifikation der Mehrzahl der HSIL-Fälle). Daher konnten weder Sensitivität noch Spezifität, sondern nur die relative Rate echt positiver Befunde, die Relative TPR und die relative Rate falsch positiver Befunde, die Relative FPR ermittelt werden. Die Relative TPR von 1,13 sprach indirekt für eine etwas höhere Sensitivität von ThinPrep™ und die ermittelte Relative FPR von 1,12 für eine etwas geringere Spezifität von ThinPrep™.

Metaanalyse zur Bewertung der Testgüte konventioneller Pap Tests

Zur Bewertung der Testgüte des konventionellen Pap Tests wurden insgesamt 84 Studien mit variierender Qualität eingeschlossen. Darin wurden Sensitivitätswerte von 6% - 99% und Spezifitätswerte von 6% – 99% berichtet. Die nach der oben beschriebenen Methode ermittelten Effectiveness Scores reichten von 1,03 (95% CI 0,78 - 1,14) an der Schwelle ASCUS/CIN 1 bis 1,29 (95% CI 1,08 – 1,50) an der Schwelle HSIL/CIN 2 - 3. Die Scores waren damit relativ niedrig, eine bessere Dis-

krimation wurde bei höheren Schwellenwerten gesehen. Wie unter dem Punkt „Metaanalyse“ beschrieben, wurden durch Einsatz verschiedener analytischer Verfahren Testgütekriterien für niedrigprävalente (10% - 19%) Screeningpopulationen berechnet. Als Gesamt-Sensitivität wurde von den Autoren 51% (95% CI 37% – 66%) angegeben, als Gesamt-Spezifität 98% (95% CI 97% – 99%). In den Berechnungen wurde auf verschiedene Störeinflüsse kontrolliert. Als größte Einflussgröße für Verzerrungen wurde dabei die Krankheitsprävalenz ermittelt: Höhere Krankheitsprävalenz war mit höherer Sensitivität und geringerer Spezifität assoziiert (mit einem größeren Effekt auf die Spezifität). Die Autoren diskutierten diesen Effekt der Prävalenz auf die grundsätzlich prävalenzunabhängigen Größen Sensitivität und Spezifität als Ausdruck eines Work-up-Bias (Systematischer Fehler, der dadurch verursacht wird, dass in einer Studiengruppe zahlenmäßig mehr inkomplett oder unkorrekt diagnostizierte Fälle enthalten sind als in der Vergleichsgruppe. Gewöhnlich tritt dieser Fehler auf, wenn Screeningpatienten mit positivem Testergebnis weiteren diagnostischen Maßnahmen unterzogen werden („Gold-Standard“), während Patienten mit negativem Testergebnis nicht diesen Maßnahmen unterzogen werden (Last 2001)).

f) Schlussfolgerungen

- Schlussfolgerung 1 (bezüglich konventioneller Pap Tests): Der konventionelle Pap Test hat eine schlechtere Fähigkeit, zwischen kranken und gesunden Personen zu unterscheiden, als bisher angenommen. Der Pap Test ist genauer an einer Schwelle HSIL+ als bei der Detektion niedriggradiger Dysplasien. Niedrigere Testschwellen resultieren in einer schlechteren Unterscheidungsfähigkeit.
- Schlussfolgerung 2 (bezüglich der Dünnschichtpräparationstechniken): Basierend auf Studien, die einen zytologischen oder histologischen Referenzstandard verwendeten, zeigte ThinPrep™ eine erhöhte Sensitivität im Vergleich zu konventionellen Pap Abstrichen. Die einzige Studie, welche die Abschätzung einer relativen FPR erlaubte, deutet auf eine etwas geringere Spezifität von ThinPrep™ gegenüber konventionellen Pap Tests hin, dieser Unterschied war jedoch statistisch nicht signifikant.
- Schlussfolgerung 3 (bezüglich computergestützter Auswertung von Pap Tests): Die Computertechnologie AutoPap™ hat gezeigt, dass sie in der Lage ist, die Rate der falsch Negativen zu senken, obwohl es schwierig ist, das Ausmaß der Reduktion der falsch Negativen unter verschiedenen Umständen präzise zu bestimmen. Ähnlich unzuverlässig ist die Abschätzung des Effekts des Rescreenings auf die Spezifität.

- Schlussfolgerung 4 (bezüglich aller neuen Technologien): Die Evidenzlage bezüglich neuer Technologien im Zervixkarzinom-Screening ist aus verschiedenen Gründen nicht ausreichend. Erstens ist nur wenig Evidenz erhältlich, aus der ein Effekt von Dünnschichtpräparationstechnik oder computerisiertem Rescreening auf die Spezifität abgeleitet werden kann. Zweitens basieren die meisten Schätzungen der Sensitivität (oder Reduktion der Rate der falsch Negativen) auf Surrogaten des Referenzstandards: der Zytologie. Standards für zytologische Referenzstandards, herausgebracht von der FDA und der Intersociety Working Group for Cytology Technologies, fordern zur Verbesserung der Validität dieser Standards ein Konsensurteil eines unabhängigen zytologischen Expertengremiums. Dieser Standard wird jedoch häufig nur bei diskrepanten Fällen eingesetzt und ebenso fehlt häufig die histologische Verifikation hochgradiger Dysplasien. Beide Mängel führen zur Überschätzung der diagnostischen Leistungsfähigkeit eines Tests.

g) Abschließende Beurteilung

Im vorliegenden HTA Report wurde erstmals eine umfassende Bewertung der Leistungsfähigkeit des konventionellen Pap Tests anhand seiner Testgütekriterien vorgenommen. Trotz der in epidemiologischen Verlaufsdaten demonstrierten positiven Effekte auf Inzidenz und Mortalität des invasiven Zervixkarzinoms seit Einführung des Zervixkarzinom-Screenings mit dem Papanicolaou-Test ergab sich eine überraschend geringe Sensitivität dieses Tests von etwa 51% bei gleichzeitig hoher Spezifität in der dazu angefertigten Metaanalyse. Alle dazu unternommenen Schritte sind von den Autoren systematisch aufbereitet und dargelegt. Bezüglich der neuen Technologien (AutoPap™, PAPNET™ und ThinPrep™) wird ausführlich auf die Probleme unzureichender Studienergebnisse und –qualität eingegangen, Verzerrungsmöglichkeiten, wie sie z.B. aus der Wahl der Stichprobe oder fehlender Verifikation von negativen Testergebnissen erwachsen, werden ausführlich diskutiert und in Schlussfolgerungen konsequent umgesetzt.

Der gesamte Report stellt eine hoch systematische Arbeit mit zugrundeliegender umfassender und ebenfalls systematischer Literaturrecherche dar. Alle Ergebnisse wurden übersichtlich präsentiert, methodische Limitationen wurden kritisch diskutiert und in den Schlussfolgerungen konsequent berücksichtigt. Die Ergebnisse des Reports zu den Testgütekriterien Sensitivität und Spezifität sind auf deutsche Verhältnisse gut anwendbar. Bezüglich aller Aussagen zu Rescreeninguntersuchungen bestehen Einschränkungen in der Übertragbarkeit, da in Deutschland kein randomisiertes zehnzehntiges Rescreening vorgeschrieben ist. Auch sind aufgrund anderer Finanzierungs- und Organisationsstruktur des US-amerikanischen Gesundheitswesens wesentliche Einflussgrößen vorhanden, die eine direkte Übertragung der Ergebnisse auf Deutschland verbieten. Gut vergleichbar sind beide Länder hinsichtlich der empfohlenen Screeningintervalle von einem Jahr.

I C 5. 2. 2 Berücksichtigter HTA-Bericht mit systematischer Bewertung aller Dünnschichtpräparationstechniken und systematischer Review über die ThinPrep™-Dünnschichtpräparationstechnik

Im folgenden Abschnitt werden Dokumente vorgestellt, die sich ausschließlich mit der Bewertung von Dünnschichtpräparationstechniken befassen.

HTA-Bericht: Payne N, Chilcoott J, McGoogan E: Liquid-based cytology in cervical screening: a rapid and systematic review. *Health Technol Assess* 2000; 4 (18).

a) Dokumenttyp und Bezugsrahmen

Bei dem hier beschriebenen Dokument handelt es sich um einen HTA-Report des National Coordinating Centre for HTA (NCCHTA, Großbritannien). Er wurde nach den Kriterien eines sogenannten Rapid Review, d.h. eines schnellen HTA-Berichts verfasst. Er ist Teil des NHS R&D HTA-Programms, das sicherstellen soll, dass qualitativ hochwertige Informationen aus der Forschung über Effektivität, Kosten und Einfluss von Gesundheitstechnologien ausgewertet werden und im Ergebnis zu einer effizienten Anwendung im Gesundheitswesen führen. Dabei werden die Projekte von einer ständigen HTA-Kommission, die von sechs Advisory Panels unterstützt wird, identifiziert und priorisiert.

Im vorliegenden Rapid Review ging es darum, den Stand der Forschung über flüssigkeitsbasierte Präparationstechniken (In der wissenschaftlichen Literatur werden die Begriffe „Thin layer“, „Monolayer“ für Dünnschichtpräparationstechniken und „Liquid-based Cytology“ für flüssigkeitsbasierte Präparationstechnik in der Zytologie synonym benutzt.) (Liquid-based Cytology, LBC) zu referieren, um im Ergebnis den Prozess der Leitlinienentwicklung durch das National Institute for Clinical Excellence (NICE) zu unterstützen.

b) Konkrete Fragestellung

Im vorliegenden Bericht sollte der Frage nachgegangen werden, wie effektiv und effizient die flüssigkeitsbasierte Zytologie verglichen mit konventionellen Pap Tests im Zervixkarzinom-Screening ist.

c) Zielpopulation

Bei der Zielpopulation handelt es sich um die gleiche, wie im bereits existierenden Screeningprogramm, d.h. es werden Frauen im Alter zwischen 20 und 64 Jahren alle 3-5 Jahre zu einem freien Screening eingeladen. Die Autoren begründeten diese Gleichsetzung mit der Annahme, dass die neuen Technologien der flüssigkeitsbasierten Präparationstechniken die konventionelle Methode ablösen werden und erwarteten darum auch keine bedeutsamen Subgruppen.

d) Methode

Recherche: Die Autoren verwendeten drei Literatursuchstrategien in Anlehnung an den AHTAC-Report: 1. Suche nach klinischer Effektivität, 2. Suche nach Kosten-Effektivität und 3. Suche nach Modelling-Studien mit generalisierter Aussage. Dafür wurden folgende Datenbanken systematisch abgefragt: Medline; EMBASE; Science Citation Index; Cochrane Library; NHS CRD: DARE, NEED and HTA; HealthSTAR; National Research Register. Die Suchstrategie wurde klar expliziert. Zu den im AHTAC-Report publizierten Studien wurde eine Zitatensuche durchgeführt. Darüber hinaus fanden Industriesubmissionen an das NICE (National Institut for Clinical Excellence) Aufnahme in das Dokument und es wurden Webseiten der HTA-Organisationen über aktuelle HTA-Projekte zum Thema kontaktiert. Die Literaturrecherche erfolgte für den Zeitraum von 1996 bis November 1999 mit einem Update im Februar 2000.

Eingeschlossen wurden alle HTA-Dokumente und sekundären Forschungsarbeiten. Primäre Forschungsarbeiten wurden aufgenommen, wenn sie über folgende Outcomeparameter berichteten: Vergleich der LBC mit konventionellen Pap Abstrichen unter Bewertung von Sensitivität und/oder Spezifität, Kategorisierung der Proben, Prozentsatz der inadäquaten oder nicht ausreichenden Abstriche, Interpretationszeiten der Proben. Ein Einschluss erfolgte jedoch nur dann, wenn numerische Daten in Tabellen klar dargestellt wurden.

Datenextraktion: Die Extraktion der Daten wurde von einem der Autoren durchgeführt. Schlüsseltabellen und Berechnungen der zusammenfassenden Tabellen wurden geprüft, indem die Daten der Studien (sofern verfügbar) in Kalkulationstabellen eingegeben wurden und die relevanten Prozentsätze nachgerechnet wurden.

e) Ergebnisse

Identifizierte Primärstudien waren teilweise bereits Bestandteil der zitierten HTA-Reports, darüber hinaus konnten keine RCT's aufgefunden werden, die Inzidenz

oder Mortalität als Outcome verwendet haben mit daraus resultierender Unsicherheit bei den Schlussfolgerungen. Weiterhin wurden keine Studien gefunden, die an typischer Screeningpopulation durchgeführt wurden und die für alle untersuchten Frauen den Goldstandard Histologie eingesetzt haben. Die Autoren schlugen alternative Verifikationsmöglichkeiten vor: (1) Studien mit zytologischem Referenzstandard eines Expertengremiums oder (2) Studien mit zusätzlichen Untersuchungen (Koloskopie oder Biopsie) an Hochrisikogruppen.

Die von den Autoren berichtete Sensitivität war bei der LBC-Methode höher oder gleich hoch wie bei konservativen Pap Tests, häufig handelte es sich in den zitierten Studien jedoch nur um kleine Fallzahlen und die Unterschiede zwischen den Methoden waren gering und/oder nicht statistisch signifikant.

Die Autoren berichteten einen über alle Studien wahrnehmbaren Trend zum Upgrading, d.h. Höherklassifizieren einer Diagnose bei den Dünnschichtpräparaten.

Die zitierten Studien waren von unterschiedlicher Größe und Qualität, ebenso die statistische Signifikanz der Unterschiede. Metaanalytische Ansätze zur Ergebnisszusammenfassung waren nicht einsetzbar, da die Populationen aufgrund unterschiedlicher Prävalenzen untereinander nicht vergleichbar waren.

Von den Autoren wurde das „Split-Sample-Design“ als methodisch ungünstig diskutiert, da es die dabei gewonnenen Dünnschichtpräparate benachteiligt (Nach Brown & Garber (1998, in Broadstock 2000) führt das Split-Sample-Design zu einer Benachteiligung der Dünnschichtverfahren mit Unterschätzung der Sensitivität der flüssigkeitsbasierten Präparationen. Von Corkill et al. (1997, in Broadstock, 2000) konnte gezeigt werden, dass wenn eine Abstrichprobe direkt in die Flüssigkeitsphiole gegeben wurde, dass die daraus gewonnenen Objektträger die gleiche Menge endozervikalen Materials enthielten, wie die bei konventionellen Abstrichen. Da beim Split-Sample-Design jedoch immer zuerst ein konventioneller Abstrich gefertigt wird und erst danach aus dem verbleibenden Material ein Dünnschichtpräparat hergestellt wird, fand sich bei diesem Design häufig ein geringerer Teil endozervikaler Zellen im Dünnschichtpräparat (AHTAC 1998)) und da es darüber hinaus im direkten Vergleich ohne Verifikation keine Aussagen über Sensitivität, Spezifität, PPV und NPV erlaubt.

Die Mehrzahl der Studien berichtete über eine höhere Anzahl unbefriedigender Objektträger (Inadequate specimens) in konventionellen Pap Tests als in Dünnschichtpräparationen. Dabei führten die Autoren an, dass die in den Studien eingesetzten Zahlen für Inadequate specimens bei konventionellen Pap Tests weit unter denen des NHS lagen und führten eine einfache zusammengefasste gepoolte Datenanalyse durch. Diese Analyse ergab etwa eine Halbierung der Rate unbefriedigender Abstriche durch Dünnschichtverfahren für das NHS.

In einigen Studien wurde der Parameter Interpretationszeit erwähnt. Scheinbar würden Screeningzeiten durch Dünnschichtverfahren abgekürzt (3 vs. 4-6 Minuten) aber gleichzeitig gäbe es Hinweise für eine erhöhte Anforderung an die Konzentration des Zytologen und darum eine schnellere Ermüdung/höhere Anzahl Pausen bei Dünnschichtverfahren.

Die Autoren wiesen darauf hin, dass für Zytopathologen eine spezielle Ausbildung in der Interpretation von Dünnschichtpräparaten erforderlich sei und in Studien eine initiale Überinterpretation der Befunde im Sinne eines Upgrading von Diagnosen dargelegt worden sei.

Die Ergebnisse des Reports wurden von uns stark gekürzt und auf unsere Forschungsfragen bezogen berichtet. Darüber hinausgehende Ergebnisse des Berichts werden nur im nachfolgenden Teil „Schlussfolgerungen“ genannt.

f) Schlussfolgerungen

Vorteile der LBC-Zytologie:

- Abnahme der Zahl inadäquater Abstriche,
- Erhöhung der Sensitivität, obwohl diese anhand vorliegender Daten schlecht zu quantifizieren sei,
- möglicherweise abnehmende Untersuchungszeiten;
- potentiell bessere Applikation eines HPV-Tests an bereits gewonnener Probe bei LBC Proben (vom National Screening Committee wird eine HPV-Status-Bestimmung in der Triage von Frauen mit milden und Borderline-Befunden als sinnvoll empfohlen),
- potentiell LBC-Objekträger in automatisierten Screeninggeräten einsetzbar.

Nachteile der LBC-Zytologie:

- keine RCT's, die Patientenbezogene Outcomeparameter vergleichen (Inzidenz,...);

- technologiebezogene erhöhte Kosten und potentielle Einspareffekte durch Reduktion der Zahl inadäquater Abstriche nicht quantifizierbar;
- individuell unterschiedlich langes Training für Laborpersonal/Zytologen erforderlich, evtl. auch für Abstrichnehmer;
- nur wenige histologisch verifizierte Sensitivitäts-Studien, auch Spezifität nicht einschätzbar, evtl. sogar schlechter für neue Methoden;
- Unklarheit über Vergleichsmöglichkeiten von LBC mit unterschiedlichen Rescreeningpraktiken zur Steigerung der Qualität der Befunde;

Laut Empfehlung des American College of Obstetricians and Gynecologists, 1998, sind Dünnschichtpräparationen nicht als Standard anzusehen, da keine populationsbezogenen Untersuchungen über Patientenbezogene Outcomeparameter existieren. Es muss verhindert werden, dass eine Ressourcenallokation stattfindet, in der bei Bemühungen zur Senkung der falsch Negativen-Rate die Anstrengungen zur Erhöhung der Teilnahmezahlen am Screeningprogramm aus den Augen verloren werden.

Bei der Frage nach einer suffizienten Erhöhung der Sensitivität einer Screeningmaßnahme mit dem Ziel der Senkung von Morbidität und Mortalität muss immer das Gesamtscreeningprogramm bewertet werden und nicht nur der einzelne Test. Viele Karzinome treten bei ungescreenten/zu selten gescreenten Frauen auf und darum ist eine Erhöhung der Teilnahme wahrscheinlich viel effektiver als eine Erhöhung der Testgüte. Obwohl in Großbritannien bereits eine recht hohe Teilnahme vorliegt, bei der eine weitere Steigerung zunächst schwierig und teuer erscheint, wird dennoch von den Autoren davon ausgegangen, dass aufgrund der großen geographischen Varianz der Screeningdichte noch Reserven vorhanden sind und damit die Möglichkeit, dass dadurch eine höhere Effektivität und Effizienz erreicht werden kann, als durch die Steigerung der Test-Sensitivität.

g) Abschließende Beurteilung

Das vorliegende Dokument stellt eine systematische Technologiebewertung der Dünnschichtpräparationstechnik dar. Alle Untersuchungsschritte wurden klar nachvollziehbar expliziert. Die Ergebnisse wurden übersichtlich präsentiert und es wurde eine kritische Diskussion über methodische Schwierigkeiten der begutachteten Studien sowie über mögliche Verzerrungen ihrer Ergebnisse geführt. Die nachfolgenden Empfehlungen folgten festgestellter Evidenz.

Metaanalyse: Bernstein SJ, Sanchez-Ramos L, and Ndubisi B.: Liquid-based cervical cytologic smear study and conventional Papanicolaou smears: A metaanalysis of prospective studies comparing cytologic diagnosis and sample adequacy. *Am J Obstet Gynecol* 2001; 185:308-317.

a) Fragestellung

In einer ersten Metaanalyse von prospektiven Studien sollten die Entdeckungsraten positiver zytologischer Befunde sowie die Raten an auswertbaren Befunden (Adequacy) zwischen konventionellen Pap Tests und Dünnschichtpräparationstechniken (ThinPrep™) verglichen werden.

b) Patienten

Die Metaanalyse der Daten bezog sich auf 533.039 Frauen, davon befanden sich 221.864 Frauen in der ThinPrep™-Gruppe, 378.659 Frauen in der konventionellen Pap Testgruppe und 67.484 Frauen wurden im Rahmen von Split-Sample-Studien untersucht, gehörten darum zu beiden Gruppen.

c) Methode

Von den Autoren wurde für den Zeitraum zwischen Januar 1990 und April 2000 in den Datenbanken Medline, PubMed und Silver Platter (sic!, alle genannten greifen auf dieselbe Datenbank zu) eine systematische Literaturrecherche durchgeführt. Von den dabei identifizierten 49 Publikationen wurden 24 ausgeschlossen, die andere Dünnschichtpräparationsverfahren einsetzten, die keinen Vergleich zwischen konventionellen und ThinPrep™-Objektträgern anstellten oder bei denen die spezifischen Outcomeparameter fehlten.

Die Autoren berichteten über den Einschluss von 24 prospektiven Studien und 1 Abstract (davon 18 aus U.S.A., 1 aus Kanada, 1 aus Japan, 2 aus Australien, 1 aus der Schweiz, 1 aus Costa Rica, 1 aus Taiwan), die folgende Einschlusskriterien erfüllten:

- Verwendung des Bethesda-Systems für die Nomenklatur zytologischer Befunde
- Untersuchung von der Adequacy der Abstrichproben sowohl für konventionelle als auch für ThinPrep™-Objektträger

- Klare Dokumentation der in Split-Sample-(SpSa) oder Direct-to-Vial-Design (DVT) erhobenen Daten

Split-Sample- und Direct-to-Vial Studien wurden separat ausgewertet, Studien über ThinPrep™ Modell Beta und Modell 2000 wurden sowohl gemeinsam als auch unabhängig ausgewertet.

Berichtete Outcomeparameter waren:

1. Häufigkeit aufgefundener Diagnosen ASCUS (hier zusammengefasst ASCUS und AGUS wegen geringer Anzahl AGUS), LSIL, HSIL;
2. Auswertbarkeit der Abstrichproben (Adequacy): adäquater Abstrich (satisfactory) definiert als Abstrich, der Plattenepithelien, Endozervikalzellen und ggf. metaplastische Zellen der Transformationszone enthält; unsatisfactory, definiert als Abstrich mit ungenügendem Anteil an Plattenepithelzellen, Abwesenheit von Endozervikalzellen, überdeckende Blut- oder Entzündungszellen, dicker Ausstrich, schlechte Fixierung oder Zytolyse.

Für jeden Outcomeparameter wurden Odds Ratio (OR) und 95% Konfidenzintervalle (95% CI) unter Verwendung von Fixed-Effect-Models (Mantel-Haenszel) und Random-Effects-Models berechnet. Da die Unterschiede zwischen beiden Methoden nicht substantiell waren, wurden nur die Ergebnisse der Fixed-Effect-Models berichtet. Ein Heterogenitätstest nach Breslow and Day wurde durchgeführt, um zu prüfen, ob die Ergebnisse der Einzelstudien zusammengefasst werden konnten.

d) Ergebnisse

Die Häufigkeit aufgefundener ASCUS-Diagnosen zeigte keinen wesentlichen Unterschied zwischen beiden Methoden.

In Bezug auf aufgefundene LSIL-Diagnosen war ThinPrep™ signifikant besser (OR 2,15 in DVT, 95% CI: 2,05; 2,26; OR 1,27 in SpSa; 95% CI 1,21; 1,32).

Eine deutliche Überlegenheit der ThinPrep™ Methode gegenüber der konventionellen Methode (OR 2,26 in DVT, 95% CI: 2,06; 2,47) zeigte sich ebenfalls, wenn die Häufigkeit aufgefundener HSIL-Diagnosen verglichen wurde, auch wenn für beide Gerätegruppen (ThinPrep™ Modell Beta und Modell 2000) getrennt ausgewertet wurde. Dieses Ergebnis fand sich auch in SpSa-Studien.

Die Auswertbarkeit der Abstriche (Adequacy) zeigte in den gepoolten Daten eine signifikante Verbesserung der ThinPrep™ Präparate über alle Studien. (OR 2,11 in DVT: 95%CI 2,07 - 2,15; OR 1,64 in SpSa: 95% CI: 1,53 - 1,76).

e) Schlussfolgerung der Autoren

Bernstein et al. schlussfolgerten, dass ThinPrep™-Präparationen zu einer deutlichen Verbesserung der Auswertbarkeit der Proben (Adequacy) geführt haben, dass ThinPrep™ signifikant höhere Entdeckungsraten von LSIL- und HSIL-Diagnosen aufwies, dass keine signifikante Veränderung bei den Entdeckungsraten von ASCUS/AGUS zu beobachten war.

Bei der von den Autoren demonstrierten Überlegenheit der ThinPrep™ Methode gegenüber konventionellen Pap Abstrichen in den hier verglichenen Outcomeparametern forderten sie weiterführende histologisch verifizierte Studien, die zur Ermittlung der Testgütekriterien Sensitivität und Spezifität erforderlich sind, da sich erst darin eine wirkliche Verbesserung des Zervixkarzinom-Screenings darstellen ließe.

f) Abschließende Bewertung

Die Ergebnisse dieser Arbeit müssen vorsichtig betrachtet werden. In allen durchgeführten Meta-Analysen wurde eine hohe Heterogenität festgestellt, was von den Autoren auf die unterschiedlichen zugrunde liegenden Bevölkerungen der einzelnen Studien („low-risk“, „high-risk“) zurückgeführt wurde. In der Arbeit fehlte jedoch eine Subgruppen-Analyse, um diesen und anderen Ursachen der Heterogenität nachzugehen. Darüber hinaus fehlten in der Arbeit ausführliche Angaben zu den eingeschlossenen Studien, die eine genauere Beurteilung der Ergebnisse ermöglichen. Die Publikation wurde in die vorliegende Arbeit aufgenommen, obwohl es sich bei den berichteten Outcomeparametern um eine Evaluation auf der Stufe technischer Machbarkeit (Flynn, 1996) handelte, was nicht den von uns definierten Ergebnisparametern entsprach. Unser Ziel war es jedoch, den besten Grad an verfügbarer Evidenz darzustellen und die Publikation war der einzige von uns ermittelte systematische Review zum Thema im Publikationszeitraum.

Tabelle 16: Zusammenfassung der berücksichtigten HTA-Reports zu klinischer Effektivität

Quelle, HTA-Organisation	Fragestellung	Informativthese	Ergebnisse, Schlussfolgerungen	Kritikpunkte der Autoren	Empfehlung der Autoren	Kommentar
McCrory et al., 1999 AHCPR, USA	<p>Determinierung der Testgüte von</p> <ol style="list-style-type: none"> 1) konventionellem Pap Test und 2) Verfahren der Dünnschichtpräparationen und 3) computerisierte Verfahren (neuronaales Netzwerk und algorithmische Klassifikation) 	<p>Qualitative Bewertung von Studien</p> <p>Quantitative Synthese: Metaanalyse der Testgüte konventioneller Pap Tests</p>	<p>Konventioneller Pap Test: Basierend auf n=84 Studien, betrug die Gesamtsensitivität des Pap Tests 51% (95% CI 37-66%); dabei Spezifität 98% (95% CI 97-99%)</p> <p>Sensitivität des Pap Tests größer für hochgradige Dysplasien</p> <p>Pap Test Genauigkeit und Prävalenz: Sensitivität höher und Spezifität geringer bei höherer Krankheitsprävalenz; Assoziation möglicherweise Ausdruck eines Work-up-Bias</p> <p>Dünnschichtpräparationstechniken: Bei kolposkopischer/histologischer Verifikation Sensitivität und Spezifität vergleichbar mit Pap Test; bei zytologischem Referenzstandard deutliche Verbesserung der Sensitivität; aus Daten keine Aussage über Spezifität der Dünnschichtpräparationen erhältlich;</p> <p>Computergestützte Auswertung mit AutoPap™: Bei kolposkopischer/histologischer Verifikation Sensitivität und Spezifität vergleichbar mit Pap Test; bei zytologischem Referenzstandard deutliche Verbesserung der Sensitivität; aus Daten keine Aussage über Spezifität von AutoPap™ erhältlich.</p>	<p>Stark variierende Qualität der untersuchten Studien</p> <p>Zu wenige Studien erfüllten Originalabschlusskriterien, darum Modifikation der Einschlusskriterien erfolgt</p> <p>Hauptausschlusskriterien: fehlende histologische/kolposkopische Referenzstandards (besonders bei neuen Technologien) und fehlende Daten für die Kreuztabelle</p> <p>Bei Einsatz zytologischer Referenzstandards häufig nicht die dafür geforderten Bedingungen (unabhängiges Expertengremium, Konsensusurteil auch bei konkordanten Testergebnissen) erfüllt</p> <p>Verzerrungen der Studienergebnisse durch "Work-up-Bias", da die meisten Studien an Frauen mit initial auffälligem Zytologiebefund vorgenommen wurden</p>	<p>Computerisierte Auswertungsverfahren (wie auch Dünnschichtpräparationsverfahren) erhöhen Effektivität bei höheren Kosten, wegen erheblicher Unsicherheit bei den Schätzern von Sensitivität und Spezifität ist klare Empfehlung jedoch problematisch.</p> <p>Weitere Studien mit Verifikation der negativen Befunde zur Bestimmung der Spezifität erforderlich.</p>	<p>Ausführlicher Bericht mit klar explizierten Methoden und Ergebnissen</p> <p>Datenextraktion durch Teams mit zunächst unabhängiger Auswertung und Konsensusprozess</p>

Quelle, HTA-Organisation	Fragestellung	Methodik der Informationssynthese	Ergebnisse, Schlussfolgerungen	Kritikpunkte der Autoren	Empfehlung der Autoren	Kommentar
AHTAC-Report, 1998	<p>Hauptfokus: Bewertung von Effektivität und Kosteneffektivität von:</p> <ol style="list-style-type: none"> 1) Geräten im Resceen- ningsersatz konventioneller Pap Abstriche 2) Geräten im Primärsceen- ningsersatz 3) Geräten zur verbesserten Präparations- technik 4) Geräten zum computerisier- ten Monitoring von manuellen Screeningpro- zessen 	<p>Qualitative Bewertung von Studien keine Meta- analyse</p>	<p>Neue Technologien insgesamt: Durch Einsatz neuer Technologien resultieren- der Zuwachs an diagnostizierten hochgradigen Dysplasien wird auf 0,05% für australische Labors geschätzt. Begrenzte Evidenz über automatische und halbautomatische Screeninggeräte, keine validen statistischen Angaben über den Vergleich mit konventionellen Methoden. Computergestützte Auswertung mit AutoPap™: Algorithmische Klassifikation: AutoPap™ zeigte erhöhte Detektionsraten gegenüber 10% randomisiertem Rescreening; ohne Relevanz für Australien wegen anderer verwendeter Re- screeningstrategien. Dünnschichtpräparationstechniken: Verringerung des Anteils nicht auswertbarer Objektträger, erhöhte Detektionsraten, hoher Übereinstimmungsgrad mit konventionellen Pap Tests, ThinPrep™ und AutoCytePREP™ wohl ähnlich, bei AutoCytePREP™ jedoch Datenlage noch schlechter, signifikante Lernkurve bei auswertenden Zytologen.</p>	<p>AutoPap™-Studien nur in geringer Zahl vorhanden und schwierig zu interpretieren. ThinPrep™-Studien zumeist in split-sample-Technik, fehlende histologische Verifikationen.</p>	<p>Die meisten Todesfälle an Gebärmutterhalskrebs gibt es bei nicht oder nicht regulär gescreeenten Frauen. Diese Fälle sind durch Erhöhung von Sensitivität und Spezifität des Screeningtests nicht beeinflussbar. Abwägung von Kosten und Ressourcenverbrauch zur Erreichung des Ziels der Senkung von Inzidenz und Mortalität des Zervixkarzinoms.</p>	<p>Report referiert den Wissensstand von 1997. Screeningprogramm Australiens differiert erheblich von Deutschland.</p>

Quelle, HTA-Organisation	Fragestellung	Methodik der Informations-synthese	Ergebnisse, Schlussfolgerungen	Kritikpunkte der Autoren	Empfehlung der Autoren	Kommentar
<p>Broadstock, M, 2000; NZHTA, New Zealand</p>	<p>Bewertung der klinischen Effektivität anhand von Sensitivität und Spezifität bei automatisierten und halbautomatischen Screeninggeräten im Vergleich zu konventionellen Pap Abstrichen</p> <p>Ermittlung der Evidenz für einen Einsatz dieser Geräte im Zervixkarzinom-Screeningprogramm Neuseelands</p>	<p>Qualitative Synthese, Evidenztabellen, keine Metaanalyse</p>	<p>Dünnschichtpräparationsverfahren: keine Evidenz für Überlegenheit neuer Technologien, da insbesondere bei hochgradigen Dysplasien keine Verbesserung der Sensitivität, wenig oder keine Information über Spezifität</p> <p>ThinPrep™: 8 Systemat. Reviews, 10 Primärstudien</p> <p>AutoCyte PREP™: 3 Systemat. Reviews, 3 Primärstudien</p> <p>Computergestützte Auswertung mit AutoPap™: Keine Evidenz für erhöhte Sensitivität oder Spezifität; keine Evidenz für erhöhte Detektionsraten bei hochgradigen Dysplasien;</p>	<p>Keine RCT's mit Outcome invasives Karzinom (weder Inzidenz- noch Mortalitätsbezogen)</p> <p>Studien hochgradig limitiert durch das Design, inadäquate Referenzstandards und inkomplette Verifikation der zytologischen Diagnosen</p> <p>Nach Auffassung der Autorin Verwendung falscher Schwellenwerte in den Studien (Schwelle zumeist auf ASCUS+ oder LSIL+ festgesetzt, die häufige Spon-tanremissionen zeigen)</p> <p>Hinweis der Autorin auf Bedeutung der Spezifität für Screeningtests: geringe Spezifität bedeutet hohe Zahl falsch Positiver mit unerwünschten Auswirkungen für Betroffene</p> <p>Die meisten Studien wurden an Hochrisikopopulationen mit eingeschränkter Vergleichbarkeit zu bevölkerungsbezogenem Screening vorgenommen.</p>	<p>Einführung neuer Technologien in Neuseeländisches Zervixkarzinom-Screeningprogramm nicht empfohlen;</p> <p>Weitere Forschung zur zuverlässigen Ermittlung der Testgütekriterien wird empfohlen</p>	<p>Bestehender hoher Qualitätsstandard in Neuseeländischen Labors ist durch neue Technologien schlecht beeinflussbar;</p> <p>Nationale Besonderheiten sind bedeutsam bei der Diskussion der Ergebnisse</p>

Quelle, HTA-Organisation	Fragestellung	Methodik der Informations-synthese	Ergebnisse, Schlussfolgerungen	Kritikpunkte der Autoren	Empfehlung der Autoren	Kommentar
Payne et al., 2000; NHS R&D HTA-Programme	Effektivität und Effizienz der flüssigkeitsbasierten Zytologie im Vergleich zu konventionellen Pap Abstrichen (Nur Dünnschichtpräparationen)	Qualitative Synthese Evidenztabelle keine Metaanalyse	<p><u>ThinPrep™</u>: Einschluss von 8 Studien mit Aussagen über Sens. und/oder Spez.</p> <p>Sensitivität konservativer <u>Pap Tests</u> in Studien von 34,5% - 85,1%, überwiegend bei LSIL+ (Ausnahme Inhorn et al., 1998, Studie an wenigen Fällen mit bekanntem Zervixkarzinom, hier Sensitivität von 93,6%) Sensitivität <u>ThinPrep™</u>: 71,4% - 95,2%, überwiegend LSIL+; (Inhorn et al. 95,7%)</p> <p>Spezifität konservativer <u>Pap Tests</u>: 36,4% - 76,9% (aus 3 der gen. 8 Studien; LSIL+ Spezifität <u>ThinPrep™</u>: 58,3% - 76,2% (3 von 8 Studien, LSIL+ Schwelle)</p> <p><u>AutoCyte PREP™</u>: Einschluss von 3 Studien mit Aussagen über Sensitivität, keine Studie zu Spezifität</p> <p>Sensitivität konservativer <u>Pap Tests</u>: 77,2% - 88,6% (überwiegend LSIL+-Schwelle) Sensitivität <u>AutoCyte PREP™</u>: 76,8% - 91,0% (überwiegend LSIL+-Schwelle)</p> <p><u>Weitere Outcomeparameter</u> Inadäquate oder Unsatisfactory Specimen: Konservativer Pap Test: 0,22% - 17,3% (23 eingeschlossene Studien); <u>ThinPrep™</u>: 0,26% - 8,5% (insgesamt 13 Studien eingeschlossen) <u>AutoCyte PREP™ (und CytoRich™)</u>: 0,0% - 3,8% (insgesamt 5 <u>AutoCyte PREP™</u> und 4 <u>CytoRich™</u> Studien eingeschlossen)</p> <p>Berichtete Sensitivitäten beider Methoden innerhalb der Studien gering verschieden, und/oder statistisch nicht signifikant, im Trend etwas höher bei Dünnschichtpräparaten; Bei geringen Studienzahlen keine Unterscheidung zwischen den Leistungen der einzelnen Gerätetypen angestrebt</p>	Keine Studien mit Outcomeparametern Inzidenz oder Mortalität des Zervixkarzinoms durchgeführt Keine RCTs an typischer Screeningpopulation mit durchgängigem Einsatz des histologischen Goldstandards bei allen Probanden Tendenz zum Diagnose-Upgrading bei Dünnschichtpräparaten;	Verbesserung der Screeningstrategie erfolgsversprechender zur Steigerung der Sensitivität des gesamten Screeningprogramms als Steigerung der Sensitivität eines einzelnen Tests Einsatz der Dünnschichtpräparationstechniken hinsichtlich der zu erwartenden Absenkung der Zahl von <u>unsatisfactory / inadequate specimen</u> von Vorteil Durchführung weiterer histologisch verifizierter Studien zur Bestimmung der Testgütekriterien der Dünnschichtpräparationstechnik erforderlich	<u>Rapid Review</u> mit klar explizierter Methodik, Empfehlungen basierend auf dem Grad aufgefundener Evidenz Nationale Besonderheiten des NHS-Screeningprogramms: Intervall 3-5 Jahre, hohe Screeningabdeckung, steigende Zahlen von <u>unsatisfactory/inadequate specimen</u> Vergleich zu deutschen Verhältnissen differenziert betrachtet werden

Tabelle 17: Systematischer Review / Metaanalyse zu Dünnschichtpräparationstechniken

Quelle	Informationsquellen	Selektionskriterien	Outcomes	Synthesismethode	Ergebnisse	Bemerkungen
Bernstein et al. 2001	Medline	Englischsprachige Artikel Prospektive Studien Bethesda System ThinPrep™ vs. Konventionelle (Nur ThinPrep™-Dünnschichtpräparationen) Split-sample oder Direct-to-vial-Design	Adequacy Befundraten	Meta-Analyse (OR fixed-effects-model) Mantel-Haenszel Subgruppen Analyse (nach Studiendesign, nach Gerät)	Studien. 49 identifizierte, 25 eingeschlossen Adequacy Direct-to-vial OR=2.15 (95%CI 2.07,2.15): ThinPrep besser Split-sample OR=1.64 (95% CI 1.53,1.76): ThinPrep besser Befundsraten ASCUS/ASGUS Direct-to-vial OR=1.03 (95%CI 0.99,1.06): Kein Unterschied Split-sample OR=1.20 (95% CI 1.13,1.27): ThinPrep besser LSIL Direct-to-vial OR=2.15 (95%CI 2.05,2.26): ThinPrep besser Split-sample OR=1.27 (95% CI 1.21,1.32): ThinPrep besser HSIL Direct-to-vial OR=2.26 (95%CI 2.06,2.47): ThinPrep besser Split-sample OR=1.09 (95% CI 1.00,1.18): kein Unterschied	Alle Meta-Analyseergebnisse zeigten eine sehr hohe Heterogenität ThinPrep lieferte adäquatere Proben Aus den Befundsraten können keine Aussagen bezüglich Sensitivität oder Spezifität getroffen werden

I C 5. 2. 3 Nicht berücksichtigte HTA-Berichte

HTA-Bericht des Canadian Coordinating Office for Health Technology Assessment (CCOHTA): Assessment of Techniques for Cervical Cancer Screening. Report 1997: 2E. National Library of Canada. May 1997. ISBN 1-895561-49-3

a) Dokumenttyp und Bezugsrahmen

Bei dem vorliegenden Dokument handelt es sich um einen HTA-Bericht zur Bewertung des Zervixkarzinom-Screenings. Er wurde vom Gesundheitsministerium einer namentlich nicht benannten kanadischen Provinz bei dem Canadian Coordinating Office for Health Technology Assessment in Auftrag gegeben und mit inhaltlichen Anfragen aus Gesundheitsministerien anderer kanadischer Provinzen zur Technologiebewertung abgeglichen. Anlass dieser Untersuchung waren die auf einem 1995 stattgefundenen Workshop (Interchange '95 (Ein "Canadian Forum to Collaborate on Cervical Cancer Screening Program Implementation Strategies", abgehalten in Ottawa vom 27.02.-01.03.1995, Nachfolgeveranstaltung des National Workshop 1989. Darin wurden Möglichkeiten eines Zervixkarzinom-Screeningprogramms identifiziert und Empfehlungen zur weiteren Verbesserung erarbeitet sowie die weitere Entwicklung eines Nationalen Zervixkarzinomnetzwerks beschlossen)) formulierten Empfehlungen zur weiteren Verbesserung des Zervixkarzinom-Screenings.

b) Konkrete Fragestellung

Dem Report wurden folgende vier Ziele vorangestellt:

1. Die Untersuchung der Effektivität des Pap Tests
2. Die Betrachtung verschiedener Strategien zur Verbesserung der Effektivität des Pap Tests
3. Der Vergleich der Kosten-Effektivität verschiedener automatisierter Rescreeningstrategien
4. Die Berücksichtigung neu auftretender Technologien.

c) Zielpopulation

Für den Report wurde keine ausdrückliche Zielpopulation benannt, hinsichtlich des Risikos, an einem Zervixkarzinom zu erkranken, wurden alle jemals sexuell aktiven Frauen als Population unter Risiko beschrieben.

d) Methode

Recherche: Eine systematische Literaturrecherche wurde in den Datenbanken Medline, Cancerlit, Hlth.Plan&Admin, EMBASE, Pascal für den Zeitraum von 1985 bis 1997 durchgeführt. Zusätzliche Recherchen wurden in DIALOG PTS Newsletter, FDA Online Database, ECRINet, CURRENT CONTENTS, Clinical Medicine und in der CCOHTA-Datenbank vorgenommen. Die zugrunde liegenden Suchstrategien wurden tabellarisch klar strukturiert berichtet, Angaben zu Zeitpunkten erfolgter Recherchen wurden nicht gemacht. Die Literatursuche wurde auf englischsprachige Publikationen und humane Studien begrenzt, weitere Ein- und Ausschlusskriterien wurden nicht expliziert. Zusätzlich wurden Informationen zu automatischen Pap-Screening-Systemen incl. Kostenschätzungen von den Herstellern per Fragebogen sowie durch persönliche Kommunikation ermittelt.

Datenextraktion: Zur Methodik der Datenextraktion wurden im vorliegenden Report keine Angaben gemacht, insbesondere wurde nicht expliziert, ob die Bewertung von einzelnen oder mehreren voneinander unabhängigen Reviewern durchgeführt wurde, welche Validitätskriterien berücksichtigt wurden und welche Studien aus welchen Gründen ausgeschlossen wurden.

e) Ergebnisse

Eine systematische Technologiebewertung fand weder für Dünnschichtpräparationen noch für automatisierte Rescreeningtechnologien statt. Eine kritische Auseinandersetzung mit methodischen Mängeln von Studien fehlte.

f) Abschließende Beurteilung

Der vorliegende Report referierte ausführlich über den Stand des damaligen (1997) Wissens zum Zervixkarzinom-Screening und bezog dabei Betrachtungen über Vor- und Nachteile verschiedener Screeningstrategien sowie Berührungspunkte zu anderen Technologien, wie z.B. den HPV Tests ein. Da zu diesem Zeitpunkt nur sehr wenige Studien über neue Technologien vorlagen, die keiner systematischen Bewertung unterzogen wurden, haben wir ihn für unseren Bericht nicht berücksichtigt.

I C 5. 3 Primärstudien

Im folgenden werden die Primärstudien dargestellt. Die wichtigsten Ergebnisse der Studien werden abschließend in Evidenztabellen zusammengefasst. Die Darstellung erfolgt getrennt für die verschiedenen Technologien (AutoPap™ und die Dünnschichtpräparationstechniken AutoCyte PREP™ und ThinPrep™).

Es wurden insgesamt 11 Studien identifiziert, davon wurden 6 Studien eingeschlossen, 5 weitere Studien wurden ausgeschlossen und sind in Tabelle 22 unter Nennung des Ausschlussgrundes aufgeführt. Eine der eingeschlossenen Studien war ein RCT's, eine weitere arbeitete mit randomisierter Reihenfolge der Probenverarbeitung im Split-Sample-Design. Es konnten keine prognostischen Studien ermittelt werden, die Aussagen über den Einfluss des Zervixkarzinom-Screenings auf die Outcomeparameter Inzidenz und/oder Mortalität zuließen. Es konnten keine Studien aus Deutschland identifiziert werden. Die Qualität der Primärstudien lag zwischen B (n=1) und D (n=4) (Vgl. Qualitätsscores I A.5.2.1.).

Tabelle 18: Übersicht über Qualitätsscores der berücksichtigten Publikationen

Technologie	Anzahl der Publikationen	Davon Anzahl der RCT's	Qualität	
			Score	Anzahl
Dünnschichtpräparationen	4	1	B	1
			B-C	1
			D	2
Computergestützte Auswertung	2	0	D	2

I C 5. 3. 1 Berücksichtigte Primärstudien

Beschreibung der Studien über computergestützte Technologien zur Auswertung von Zervixabstrichen auf der Basis algorithmischer Klassifikation (AutoPap™)

In Folge der in I A.4.3. beschriebenen Literaturrecherche wurden nach Besichtigung der Abstracts zwei Veröffentlichungen identifiziert, die sich auf oben genannte Technologie in der computergestützten Auswertung von Zervixabstrichen bezogen. Die Charakteristika und Ergebnisse der Publikationen sind tabellarisch zusammengefasst (siehe Tabelle 15).

Alasio LM, Alphandery C, Grassi P, Ruggeri M, De Palo G and Pilotti S: Performance of the AutoPap Primary Screening System in the Detection of High-Risk Cases in Cervicovaginal Smears. *Acta Cytologica* 2001 ; 45; No.5: 704 – 708.

Fragestellung: In der Studie sollte die Leistung des AutoPap™-Gerätes als Primärscreener bei der Erkennung hochgradiger Dysplasien und invasiver Karzinome des Gebärmutterhalses untersucht werden.

Patienten: In die Studie aufgenommen wurden 14.779 fortlaufend gewonnene konventionelle Pap Abstriche. Davon stammten 4.715 Abstriche aus einer der Einrichtung angeschlossenen Ambulanz-Abteilung (Hochrisikopatientinnen), die übrigen 10.064 stammten von verschiedenen Screeningzentren (niedriges Risiko).

Methode: Prospektive Studie, die an einer einzelnen Einrichtung durchgeführt wurde: am Instituto Nazionale per lo Studio e la Cura die Tumori, Milan, Italy. Alle Pap Abstriche wurden zunächst manuell von Zytologen gescreent und nach dem Bethesda-System klassifiziert, alle schweren Dysplasien wurden vom Leitenden Zytologen erneut begutachtet und ein Teil der hochgradigen Veränderungen wurde anschließend bioptisch kontrolliert. Nach den Europäischen Richtlinien für interne Qualitätskontrollen erfolgte für die primär negativ befundeten Objektträger ein partielles Rescreening. Von April bis Dezember 1998 wurden alle 14.779 Objektträger mit AutoPap™ verarbeitet. Davon wurden 13.261 (90%) in einer der folgenden Kategorien ausgegeben: 10.349 als „Review“ und 2.912 als „No Further Review“. Die restlichen 1.518 (10%) Objektträger konnten aus technischen Gründen nicht vom Gerät bearbeitet werden (z.B. durch Verschmutzung oder Luftpfeitschlüsse) und wurden als „Process Review / Rerun“ ausgegeben. Die als „Review“ ausgegebenen Objektträger wurden manuell gescreent und alle pathologischen Befunde wurden einem Pathologen zur abschließenden Begutachtung vorgelegt. Diejenigen Abstriche, die anschließend mittels Biopsie auf ihren Befund kontrolliert wurden, wurden mit der vom Gerät ausgegebenen Rangfolge verglichen.

Ergebnisse: Beim manuellen Screeningprozess wurden 10.071 (97%) normale Abstrichbefunde ermittelt, atypische Zellen fanden sich in 278 (3%) aller Abstriche, davon waren 70 Abstriche hochgradige Dysplasien und Karzinome (HSIL+-Befunde). Von AutoPap™ wurden 63 / 70 Abstrichen HSIL+ (90%) in den obersten beiden Quintilen ausgegeben, 220 / 278 aller positiver Abstriche (80%) wurden gleichfalls in den obersten beiden Quintilen eingeordnet. Von 38 Abstrichen wurde eine histologische Diagnose gewonnen, dabei stimmten die histologischen Diagnosen und die manuell zytologisch gewonnenen Diagnosen zu 100% überein. Die HSIL+-Befunde wurden von AutoPap zu 100% in den Kategorien 1 und 2, d.h. in den obersten beiden Quintilen ausgegeben. Unter den vom Gerät als „No Further Review“ ausgege-

benen Befunden befanden sich keine im manuellen Screening identifizierten hochgradigen Dysplasien. Von 2.917 Abstrichen wurden 7 Unterklassifikationen beobachtet.

Schlussfolgerung der Autoren: Die Autoren kamen zu der Schlussfolgerung, dass AutoPap™ in der Erkennung schwerer Dysplasien effektiv ist und dass sein Einsatz zu einer signifikanten Reduktion falsch negativer Befunde führen sollte.

Abschließende Bewertung: In der Studie wurden zunächst alle Pap Abstriche manuell gescreent und in einem weiteren Schritt einem automatischen Primärscreening durch AutoPap™ unterzogen. Dabei wurde nicht berichtet, ob die Untersucher für das vorangehende Ergebnis des manuellen Screenings verblindet waren. Als Referenzstandard wurde der von der Intersociety Working Group for Cytology Technologies (1997) vorgeschlagene Standard einer mindestens 50%-igen histologischen Kontrolle der HSIL+-Befunde eingehalten. Über die Auswahl der 38 von insgesamt 70 HSIL+-Befunden zur histologischen Verifikation wurde nichts berichtet, auch wurde nichts über die Vorgehensweise bei diskordanten Befunden beider Untersuchungsmethoden expliziert. Ein weiterer Bestandteil der Empfehlungen der Intersociety Working Group for Cytology Technologies (1997) war jedoch das Erzielen von Konsensusdiagnosen durch ein unabhängiges Expertengremium bei Erhebung diskordanter Untersuchungsbefunde.

Bei somit unvollständigem Referenzstandard müssen die vorgestellten Ergebnisse über Testgütekriterien kritisch betrachtet werden. Darüber hinaus kann eine untersucherabhängige Verzerrung der Ergebnisse bei eventuell fehlender Verblindung nicht ausgeschlossen werden.

Renshaw AA, Lezon KM, and Wilbur DC: The Human False-Negative rate of Rescreening Pap Tests. Measured in a Two-Arm Prospective Clinical Trial. *Cancer Cytopathology* 2001; 93: 106-110.

Fragestellung: In der vorliegenden Studie sollte die Effektivität des routinemäßig erfolgenden Rescreenings (10%-iges randomisiertes Rescreening) untersucht werden. Von den Autoren wurden die Raten falsch negativer / falsch positiver Befunde des manuellen Rescreenings unter Einsatz eines analytischen Instruments (AutoPap System™) ermittelt.

Patienten: In die Studie wurden 25.124 Pap Abstriche aus 5 kommerziellen Laboratorien eingeschlossen.

Methoden: Prospektive zweiarmige klinische Studie: „Routinepraxis“-Arm: Alle Nicht-Hochrisiko-Objektträger wurden zunächst manuell von einem Zytologen gescreent und einem 10%-igen randomisierten Rescreening zur Qualitätskontrolle unterworfen. Die Befunde wurden als pathologisch, gemäß der Bethesda-Klassifikation, als „*Within Normal Limits (WNL)*“ oder „*unsatisfactory*“ ausgegeben.

„AutoPap™“-Arm: Alle Nicht-Hochrisiko-Objektträger wurden vom Gerät eingelesen und entsprechend dem dabei ermittelten Score kategorial ausgegeben: „*No Further Review*“ (25% aller erfolgreich mit AutoPap™ verarbeiteten Objektträger) war die Befundkategorie, die mit WNL gleichgesetzt wurde und keiner weiteren Begutachtung unterzogen wurde. Als „*Review*“ ausgegebene Objektträger (75% aller von AutoPap™ erfolgreich verarbeiteten Objektträger) wurden vom Zytologen unter Kenntnis der vom Gerät ermittelten Quintilzugehörigkeit (Die vom Gerät berechneten Scores werden in eine Rangfolge gebracht, von der geringsten bis zur höchsten Wahrscheinlichkeit, einen pathologischen Befund zu enthalten. Aus diesen einzelnen Rangplätzen werden Gruppen gebildet, deren Grenzen durch Quintile bestimmt werden. Siehe auch Kapitel C 2.3.1. Technische Beschreibung der Testverfahren.) manuell durchgemustert. Die als „*Quality Control Review*“ ermittelten Objektträger (15% aller erfolgreich mit AutoPap™ verarbeiteten Objektträger), d.h. der Teil der „*Review*“ Objektträger mit der höchsten Wahrscheinlichkeit, einen pathologischen Befund zu enthalten, obgleich beim Primärscreening kein auffälliger Befund erhoben wurde, wurde in der Studie einem zweiten manuellen Rescreening unterzogen. Die Befundklassifikation wurde nach identischem System wie im anderen Studienarm vorgenommen.

Alle Untersucher waren für die Untersuchungsergebnisse des anderen Studienarms verblindet. Die Studie erfolgte unter Standardbedingungen des Labors. Für alle positiven zytologischen Befunde, die in den unterschiedlichen Studienarmen diskordante Ergebnisse aufwiesen, wurden in einem mehrstufigen Konsensusprozess Diagnosen festgelegt.

Ergebnisse: Von den 25.124 in die Studie eingeschlossenen Objektträgern wurden 171 (0,7%) als „*unsatisfactory*“ und 23.556 (93,8%) als „*WNL*“ beurteilt. Von den verbleibenden 1.397 pathologischen Befunden waren 998 (4,0%) ASCUS, 51 (0,2%) AGUS, 278 (1,1%) LSIL, 67 (0,2%) HSIL, 1 (0,004%) Adenocarcinoma in situ und 2 (0,008%) invasive Tumoren.

Von den Autoren wurden für gesondert ausgewiesene Schwellenwerte (ASCUS+-Befunde und LSIL+-Befunde) die Raten falsch Negativer im Rescreening berichtet. Die Raten falsch Negativer bei der klinisch bedeutsamen Schwelle HSIL+ wurden nicht mitgeteilt. Die darüber hinaus kalkulierten Werte für ein Primärscreening wurden in unsere Auswertung nicht aufgenommen, da im Studiendesign kein gesonder-

ter Ansatz für eine Primärscreeninguntersuchung expliziert wurde. (Siehe auch Kapitel I A.6. Diskussion - Effektivität.)

Berechnet wurden:

$$\text{FNR} = \frac{\text{Anzahl der falsch negativ befundeten Objektträger im Studienarm}}{\text{Anzahl der richtig positiven Befunde im Studienarm}}$$

$$\text{FPR} = \frac{\text{Anzahl der falsch positiv befundeten Objektträger im Studienarm}}{\text{Anzahl der richtig negativen Befunde im Studienarm}}$$

Die Autoren berichteten von verschiedenen Subpopulationen (z.B. durch wechselnden Anteil „*unsatisfactory*“ variierende Zahlen in beiden Studienarmen), deren Bezug zur Gesamtstichprobe nicht dargelegt wurde. Daraus ergaben sich scheinbar widersprüchliche Zellbelegungen für die Vier-Felder-Matrix, die somit nicht vervollständigt werden konnte.

Schlussfolgerung der Autoren: Die Autoren schlussfolgerten aus ihren Ergebnissen, dass das routinemäßige 10%-ige randomisierte Rescreening eine ineffektive Methode ist, um die Rate der falsch negativen Befunde zu senken. Als wesentlich erfolgversprechender sehen sie die Möglichkeiten des Rescreeningprozesses mit AutoPap™, diese FNR zu senken, auch wenn die Rate falsch positiver Befunde dabei ansteigen würde.

Abschließende Bewertung: Die vorliegende Studie wurde über AutoPap™ im Rescreeningeinsatz durchgeführt und mit der Wirksamkeit 10%-ig randomisierten Rescreenings verglichen. Bereits bei (Broadstock, 2000) wurden verschiedene Arbeiten zitiert, die auf die Ineffektivität randomisierten 10%-igen Rescreenings zur Senkung der Rate falsch Negativer hinwiesen. Gegenwärtig werden auch in Deutschland Diskussionen über effektivere Rescreeningpraktiken als das 10%-ige Rescreening geführt. So schlugen Ruffing-Kullmann et al. (2001) das Rapid Rescreening als wirksame Alternative vor und beriefen sich dabei auch auf zwei Metaanalysen (Arbyn & Schenk, 1999 und 2000), in denen die Überlegenheit des Rapid Rescreening dargelegt werden konnte.

Tabelle 19: Evidenztabelle der Studien über computergesteuerte Technologien zur Auswertung von Zervikalabstrichen

Quelle	Fragestellung / Design	Technologie	Studienpopulation / Prävalenzangaben	Verwendeter Referenzstandard	Untersucher	Klassifikationssystem	Schwellenwert	Testgütekriterien	Weitere Ergebnisse	Bemerkungen	Studienqualität
Alasio et al., 2001	Leistung von AutoPap™ als Primärscreener bei der Detektion von hochgradigen Dysplasien und Zervixkarzinomen	AutoPap™ vs. Manuelles Screening von konventionellen Pap Abstrichen	Von 14.779 fortlaufend gewonnenen Abstrichproben von Hochrisiko-Patientinnen eigener Ambulanz, 10.064 Abstriche aus externen Screeningzentren (niedriges Risiko)	Zytologischer Referenzstandard, jedoch keine Konsensusdiagnose	k.A. über Unabhängigkeit k.A. über Verblindung	Be-the-das-System	HSIL +	[Aus Daten der Studie pro Forma ermittelt:] Sensitivität: 97% Spezifität: 40% PPV: 84% NPV: 80% [Bei fehlender kompletter histologischer Verifikation Se und Sp jedoch nicht verwendbar]	90% der HSIL+- Befunde wurden von AutoPap im Ranking in Kategorie 1 und 2 gesetzt,	Bei der Darstellung der Korrelation zwischen histologischen Diagnosen und manuellem Screening fällt eine 100%-ige Übereinstimmung auf, die Auswahl der 38 Fälle aus 70 HSIL-Befunden zur histologischen Abklärung wird nicht berichtet.	D (Inadäquater Referenzstandard, ungenügende Transparenz bei der Berichtsqualität)

Quelle	Fragestellung / Design	Technologie	Studienpopulation / Prävalenzangaben	Verwendeter Referenzstandard	Untersucher	Klassifikationssystem	Schwellenwert	Testgütekriterien	Weitere Ergebnisse	Bemerkungen	Studienqualität
Renshaw et al., 2000	Ermittlung der Raten falsch Negativer bei randomisiertem Rescreening und bei AutoPap Rescreening; Prospektive klinische Studie, nicht randomisiert	10%-iges randomisiertes manuelles Rescreening (konv.) vs. AutoPap Rescreening (AP)	25.124 eingeschlossene Objektträger aus 5 kommerziellen Laboratorien; keine weiteren Angaben über Krankheitsprävalenzen.	Zytologisches Konsensusurteil bei diskordanten Befunden	Verblindung der Untersucher gegenüber Untersucherergebnissen des anderen Studienarmes; AutoPap Rescreening unter Kenntnis des Ranking	Be-the-system	ASCUS + LSIL+	Sensitivität: k.A. Spezifität: k.A. PPV: k.A. NPV: k.A. Aus Daten der Studie nicht zu ermitteln	FNR (konv) = 73,2 FNR(AP) = 34,4 FNR (konv) = 100 FNR(AP) = 58,8	Die Studie übermittelt gleichzeitig Kalkulationen über AutoPap als Primärscreener, die bei bestehenden dem Design als Rescreeningstudie hier jedoch nicht in Betracht gezogen wurden.	D (Inadäquater Referenzstandard, da keine historische logische Verifikation eines signifikanten Anteils von hochgradig positiven Befunden erfolgte)

Beschreibung der Primären Studien über Dünnschichtpräparationstechniken (Thin-Prep™ und AutoCyte PREP™)

In Folge der in Kapitel A.4.3. beschriebenen Literaturrecherche wurden nach Auswertung der Abstracts 9 Veröffentlichungen identifiziert, die Untersuchungen zur Evaluation von Dünnschichtpräparationstechniken anstellten. Fünf dieser Publikationen entsprachen nicht den in Kapitel A.4.4. beschriebenen Auswahlkriterien. Sie sind in der Tabelle 17 mit ihrem Ausschlussgrund dargestellt.

Weitere vier Veröffentlichungen entsprachen den o. g. Auswahlkriterien und werden im folgenden diskutiert. Die Charakteristika und Ergebnisse der Publikationen sind tabellarisch zusammengefasst.

Bergeron C, Bishop J, Lemarie A, Cas F, Ayivi J, Huynh B, and Barrasso R: Accuracy of Thin-Layer Cytology in Patients Undergoing Cervical Cone Biopsy. *Acta Cytologica* 2001; 45; 519-524.

Fragestellung: In der Studie sollten Testgütekriterien von konventionellen und dünnschichtpräparierten Zervixabstrichen verglichen werden.

Patientinnen: Insgesamt nahmen 500 fortlaufend eingeschlossene Patientinnen einer Kolposkopieklinik, die zur Konisation vorgesehen waren, an der Untersuchung teil. 350 von ihnen (70%) hatten eine bioptisch gesicherte hochgradige Dysplasie des Gebärmutterhalses, 150 (30%) hatten einen diskrepanten Befund in Zytologie und Kolposkopie.

Methode: Prospektive, zweiarmige, hinsichtlich der Reihenfolge der Probenverarbeitung randomisierte klinische Split-Sample-Studie: die Patientinnen wurden gynäkologisch untersucht, ihnen wurden mit Cervexbrush™ (Fa. Rovers, Niederlande) zwei Abstrichproben entnommen, anschließend wurden sie elektrokaustisch konisiert.

Von jeder Frau wurden zwei Objektträger angefertigt: ein konventioneller, der abgestrichen und umgehend mit einem Fixierspray besprüht wurde und ein via Dünnschichtpräparationstechnik angefertigter, wobei die Abstrichprobe in die dafür vorgesehene CytoRich™-Phiole (Tri Path Imaging Inc., USA) gegeben und im Labor weiterverarbeitet wurde. Die Reihenfolge, ob zunächst ein konventioneller Abstrich angefertigt wurde oder zuerst eine Phiole zur Dünnschichtpräparation beschickt wurde, wurde randomisiert festgelegt.

Die konventionellen Abstrichproben wurden im Routineablauf des Labors *Pasteur Cerba* weiterverarbeitet und verblindet unter den Zytoassistenten zum Screening verteilt. Aus der CytoRich™-Fixierflüssigkeit wurden Objektträger in Dünnschichttechnik in einem TriPath Labor präpariert und zur Auswertung an das Pasteur Cerba zurückgesandt, wo sie von drei freiwilligen, speziell ausgebildeten Zytoassistenten gescreent wurden. Es wurde sichergestellt, dass keiner der Untersucher der Dünnschichtpräparate den gleichen Patienten wie in der konventionellen Diagnostik erhielt. Alle zytologischen Befunde wurden von einem zweiten unabhängigen Zytoassistenten ohne Kenntnis der vorbestehenden Befunde befundet und nach dem Bethesda-System ausgegeben.

Referenz-Test: Bei allen Patientinnen wurden Gewebeproben des bei der Konisation entfernten Gebärmutterhalses entnommen und histopathologisch analysiert.

Die ermittelten Zielgrößen waren Sensitivität, Spezifität, positiver und negativer prädiktiver Wert.

Ergebnisse: Von den Autoren sind die oben genannten Testgütekriterien für die Erkennung von hochgradigen CIN bei dem Schwellenwert HSIL+ berichtet worden (siehe Tabelle 20). Nach angegebenen Daten wurden die Testgütekriterien für die übrigen Schwellenwerte kalkuliert. Die Autoren berichteten von leicht erhöhter Sensitivität der Dünnschichtpräparate gegenüber konventionellen Abstrichen, der Unterschied erwies sich jedoch als nicht statistisch signifikant. Nach Kontrolle für die Reihenfolge der Probenverarbeitung zeigte sich ein statistisch signifikanter Vorteil der Dünnschichtpräparate, wenn diese zuerst aus dem Probenmaterial gewonnen wurden. Die Spezifität war in der Studie für beide Methoden niedrig (siehe Tabelle 21), was im Zusammenhang mit einer Screeninguntersuchung von den Autoren als nachteilig angesehen wurde. Wurde zuerst der konventionelle Abstrich gewonnen, war die Spezifität der herkömmlichen Methode signifikant größer als bei der Dünnschichtpräparation.

Die *Adequacy* war signifikant besser bei Dünnschichtpräparaten als in der konventionellen Methode.

Schlussfolgerung der Autoren: Die Autoren wiesen auf die Reihenfolge der Probenverarbeitung im Rahmen des Split-Sample-Designs als benachteiligenden Faktor für die in zweiter Linie verarbeiteten Dünnschichtpräparate hin. Sie betonten, dass die ermittelten Sensitivitäten und Spezifitäten nicht an einer Screeningpopulation erhoben wurden, sondern an Patientinnen einer Kolposkopieklinik mit vormals positiven Befunden. Eine Übertragbarkeit ihrer Ergebnisse auf andere Settings sei somit nicht gewährleistet.

Abschließende Bewertung: Die in die Studie eingeschlossenen Patientinnen entstammten einem speziellen Hochrisikokollektiv. Die Ergebnisse der berichteten Testgütekriterien sind auf eine niedrigprävalente Screeningpopulation nicht übertragbar (Inwieweit die primär nicht prävalenzabhängigen Größen Sensitivität und Spezifität in Settings mit höherer Prävalenz einer Verzerrung unterliegen, wird im Diskussteil näher erläutert). Das Studiendesign lieferte jedoch einen interessanten Ansatz zur Methodik von Split-Sample-Studien: unter Beibehaltung des intraindividuellen Vergleichs von Stichproben wurde sichergestellt, dass beide Untersuchungen an der gleichen Patientengruppe vorgenommen wurden, andererseits wurde der Nachteil des Split-Sample-Designs für die Dünnschichtpräparationstechnik durch eine Randomisation der Reihenfolge der Präparation ausgeglichen. Bei der Berechnung der Sensitivität und Spezifität wurden die inadäquaten Präparationen einbezogen. Da im Praxisalltag inadäquate Präparate zu einer Wiederholungsuntersuchung führen, halten wir ihren Einschluss in die Berechnung (als FP bzw. FN) für nicht sinnvoll. Die Nicht-Berücksichtigung inadäquater Präparate würde zu geringeren Unterschieden zwischen konventioneller und Dünnschichtpräparationstechnik führen. Eine weitere ausführliche Diskussion ist im Kapitel A.6. dargestellt.

Jensen ML, Fursted PB, Svanholm H. Sammenligning af monolagsprover og konventionelle „smears“. *Ugeskr Laeger* 2001;163:1270-5.

Fragestellung: In der Studie sollte die Qualität der Präparate mit AutoCyte PREP™ mit der von konventionellen Pap Tests verglichen werden. Neben der Adequacy wurden auch Befundraten, Sensitivität und benötigte Zeit verglichen.

Patientinnen: 1701 Frauen einer dänischen Kommune, die am Screeningprogramm teilnahmen.

Methode: Split-sample. Die positiven Befunde wurden histologisch weiter untersucht (Referenzstandard Portiobiopsie bzw. Konisierung). Es ist nicht zu beurteilen, ob die Untersucher der Histologie für die Ergebnisse des Abstriches verblindet waren. Die Proben wurden mittels einer Kombibürste entnommen. Die Präparation erfolgte entweder konventionell oder mit AutoCytePREP™.

Ergebnisse: Die Qualität der Präparate war mit AutoCytePREP™ besser, 0,1% der Präparate waren nicht zu beurteilen im Vergleich zu 1,8% von den konventionellen ($p < 0.001$). 11% der AutoCyte PREP™-Präparate waren schwer zu beurteilen (SBLB) gegen 18% der konventionellen. In 90% der Fälle waren die Ergebnisse beider Methoden konkordant. Die Autoren ermittelten im Vergleich mit dem Referenzstandard eine Sensitivität von 95% bei AutoCytePREP™ und von 94% bei konventionellen Pap Abstrichen. Darüber hinaus fanden die Autoren eine Zeitersparnis von durchschnittlich 2 Minuten je untersuchtem Objektträger mit AutoCytePREP™.

Schlussfolgerungen der Autoren: Die Technologie ermöglicht eine Erhöhung der Qualität der Abstriche und eine Zeitersparnis bei der Befundung.

Abschließende Beurteilung: Da nur die positiven Befunde histologisch verifiziert wurden, war die Angabe der Sensitivität nicht zulässig, es hätte allenfalls eine Berechnung der FNR erfolgen dürfen. Aus den Angaben der Studie war es nicht möglich, eine Vierfelder-Tafel aufzustellen, so dass eine nachträgliche Berechnung der FNR unsererseits nicht möglich war. Das ist insbesondere darum bedauerlich, da die hier untersuchte Patientinnen-Gruppe eine „normale“ Bevölkerung darstellt, das heißt, Verzerrungen durch den alleinigen Einschluss von Hochrisiko Patientinnen sind nicht zu erwarten gewesen wären.

Obwegeser JH, and Brack S: Does Liquid Based Technology Really Improve Detection of Cervical Neoplasia? A Prospective, Randomized Trial Comparing the ThinPrep Pap Test with the Conventional Pap Test, Including Follow up of HSIL Cases *Acta Cytologica* 2001: 45; 709-714.

Fragestellung: Primäres Ziel dieser Studie war ein Vergleich von Sensitivität, Spezifität und Beurteilbarkeit der Abstrichproben (Adequacy) zwischen konventionellen Pap Tests und ThinPrep™ Dünnschichtpräparationen an einer Population geringen Erkrankungsrisikos mit Follow-up-Beobachtung der HSIL-Fälle.

Patientinnen: Alle Patientinnen, die 15 private gynäkologische Praxen im Zeitraum zwischen Mitte Juli und Ende September 1998 (10 Wochen) zur Pap Abstrichentnahme aufsuchten, wurden randomisiert für einen entweder konventionellen oder ThinPrep™ Abstrich in die Studie aufgenommen. Insgesamt wurden 1.999 Patientinnen eingeschlossen, davon 1.002 in die konventionelle Gruppe (CV), 997 in die ThinPrep™-Gruppe (TP). Die Randomisierungszuweisung wurde dem zytologischen Fallbericht zugeordnet und in einem versiegelten Umschlag verwahrt, dessen Randomisationsnummer eine spätere Zuordnung ermöglichte. Patientinnen mit vormals positiven Pap Testbefunden wurden in die Studie eingeschlossen: 48 in TP (davon 14 LSIL, 10 HSIL, 22 ASCUS/AGUS, 2 unsatisfactory) und 50 in CV (24 LSIL, 5 HSIL, 20 ASCUS/AGUS, 1 unsatisfactory). Es wurde eine Vergleichbarkeit beider Gruppen hinsichtlich der Altersverteilung gezeigt.

Methode: Prospektive Studie mit randomisiert kontrolliertem klinischen Design (RCT): alle Frauen wurden gynäkologisch untersucht, Schleim- und Zelldetritusauflagerungen wurden mittels eines Baumwolltupfers aus dem Zervikalkanal entfernt und es erfolgte eine Abstrichentnahme unter kolposkopischer Kontrolle. Für die TP-Gruppe wurden die von der Firma Cytec Corp.® empfohlenen Abstrichentnahmeinstrumente Cervex-Brush™ (Rovers Medical Devices®, B.V., Oss, NL) oder Oribrush™ (Orifice Medical A.B.®, Ystad, Schweden) verwendet, kombiniert mit

einem Plastikspatel. Die Instrumente wurden unmittelbar nach Abstrichentnahme in der PreservCyt™ Lösung ausgeschüttelt und im Labor unter Einsatz des ThinPrep™ (Cytoc Corp.®) zu Dünnschichtpräparaten weiterverarbeitet. In der CV-Gruppe wurde der Abstrich mittels eines Szalay™-Spatels (C.S.M. Graf & Co., Steinach, Schweiz) entnommen und unmittelbar nach Entnahme ausgestrichen und in 96%-iger Alkohol-lösung fixiert. Beide Arten von Präparaten wurden nach der üblichen Papanicolaou-Färbung im Labor gefärbt. TP-Objektträger wurden von einem in Dünnschichtpräparateauswertungen in der Firma Cytoc Corp. ausgebildeten Zytoassistenten, CV-Objektträger von drei anderen, langjährig erfahrenen Zytoassistenten ausgewertet. Alle positiven Befunde wurden dem Direktor des Laboratoriums vorgelegt. Follow-up und Referenzstandard: alle HSIL-Fälle hatten ein Follow-up nach 12-15 Monaten mit entweder histologischer (im Rahmen von Konisation oder Hysterektomie) oder zytologischer Begutachtung. Die histologische Beurteilung erfolgte durch unabhängige Pathologen. Die Follow-up-Untersuchungen der LSIL- und ASCUS/AGUS-Fälle standen zum Publikationszeitpunkt noch aus.

Ergebnisse: Die zytologischen Diagnosen unterschieden sich in TP- und CV-Gruppe nicht signifikant. Die Raten unsatisfactory und SBLB (bestätigt durch Spezialisten eines Cytoc®-Labors) waren in der TP-Gruppe signifikant höher als in der CV-Gruppe und waren zumeist verursacht durch einen unzureichenden Zellgehalt. In der TP-Gruppe wurde eine 91%-ige Übereinstimmung zytologischer und histologischer Diagnosen bei HSIL-Fällen gefunden, bei der CV-Gruppe eine 100%-ige, sofern histologische Diagnosen vorlagen. Sensitivität und Spezifität wurden nicht berichtet, aus den mitgeteilten Befunden ließen sich keine Vierfeldertafeln zu deren nachträglicher Berechnung ermitteln.

Schlussfolgerung der Autoren: Die Studienergebnisse differierten deutlich von anderen direct-to-vial Studien und wurden von den Autoren mit dem Gebrauch eines adäquaten Abstrichentnahmeinstruments unter kolposkopischer Kontrolle in der CV-Gruppe begründet. Die Autoren empfahlen den Verzicht auf eine Abstrichentnahme, wenn eine Darstellung des Zervikalkanals nicht möglich sei. Außerdem wiesen sie darauf hin, dass bei Beachtung der Basisregeln für eine Abstrichentnahme (z.B. Entfernen von Schleim und Detritus aus dem Zervikalkanal) auch in anderen Studien die Beurteilbarkeit konventioneller Objektträger ebenso gut oder besser als für Dünnschichtpräparate ausfiel. Sie erinnerten an die Aufgabe eines Labors, den Klinikern Rückmeldungen über ihre Abstrichqualität zu geben.

Abschließende Bewertung: In dieser Studie, der ein randomisiertes klinisches Design zugrunde lag und bei der ein histologischer Referenzstandard für hochgradig dysplastische Befunde eingesetzt wurde, wurde gezeigt, dass sich bei Verwendung adäquater Abstrichentnahmetechniken und –instrumente keine Vorteile in der Beurteilbarkeit von Dünnschichtpräparationen nachweisen ließen. Abweichend von anderen Studienergebnissen, in denen über eine Senkung der Rate unsatisfactory und SBLB-

Abstriche berichtet wurden, fanden sich in der vorliegenden Publikation in der ThinPrep™-Gruppe weniger gut auswertbare Befunde aufgrund eines verminderten Zellgehalts. Die verminderte Übereinstimmung von HSIL-Diagnosen in der TP-Gruppe gegenüber der CV-Gruppe ist auffällig. Es wurde jedoch nichts über den Erfahrungsstand des auswertenden Zytoassistenten mitgeteilt außer seiner Teilnahme am Trainingsprogramm.

Die in der Eingangsfragestellung angekündigten Ergebnisparameter Sensitivität und Spezifität wurden leider nicht berichtet, auch ließen die publizierten Daten keine Ermittlung einer Vier-Felder-Matrix mit nachträglicher Berechnung der Testgütekriterien zu.

Park A, Lee SN, Chae SW, Park KH, Kim JW, and Lee HP: Comparing the Accuracy of ThinPrep Pap Tests and Conventional Papanicolaou Smears on the Basis of the Histologic Diagnosis. A Clinical Study of Women with Cervical Abnormalities. *Acta Cytologica* 2001; 45; 525-531.

Fragestellung: Primäres Ziel dieser Studie war die Untersuchung der Effektivität von ThinPrep™ Pap Tests im Vergleich zu konventionellen Pap Tests. Anhand der vorliegenden Studie wurden Testgütekriterien kalkuliert wie Sensitivität, Spezifität, positive und negative Vorhersagewahrscheinlichkeit.

Patientinnen: Insgesamt wurden in die Studie 483 Patientinnen einer gynäko-onkologischen Spezialambulanz des Universitätskrankenhauses Seoul aufgenommen. Bei allen Frauen lag ein vormals positiver konventioneller Pap Testbefund vor.

Methode: In einer prospektiven Split-Sample-Studie wurden Abstrichproben unter Verwendung einer Endozervixbürste (Wallach Surgical Devices, Inc., Orange, Connecticut, USA) in üblichem Split-Sample-Vorgehen (Das Abstrichentnahmeinstrument wird zunächst auf einem Glasobjektträger beidseitig ausgestrichen, der dann für eine weitere konventionelle Bearbeitung fixiert wird und im Labor einer Pap-Färbung unterzogen wird. Danach wird das Abstrichentnahmeinstrument in einer Fixierlösung (PreservCyt™ von Cytoc Corp.) ausgeschwenkt. Die Phirole mit der Zellsuspension wird dann im Labor der weiteren Dünnschichtpräparation für das AutoCyte™ Verfahren unterzogen und nach abgeschlossener Präparation ebenfalls mit der Pap-Färbung gefärbt.) gewonnen. Die zytologische Auswertung erfolgte verblindet seitens der Untersucher für das jeweils andere Resultat; die ThinPrep-Auswerter hatten ein Training bei Cytoc Inc. absolviert, von längerfristigen Erfahrungen im Umgang mit der neuen Technologie wurde nichts berichtet. Unter Kenntnis der zytologischen Befunde wurde von einem Kliniker bei insgesamt 158 Patientinnen eine Biopsie bei Vorliegen sichtbarer Veränderungen vorgenommen. Von diesen 158 histologischen Kontrollen wurden von den Autoren 18 Biopsien von der Berechnung der Testgütekriterien

ausgeschlossen, die einen unauffälligen histologischen Befund zeigten, in der zytologischen Diagnostik jedoch Dysplasien (LSIL+) aufwiesen. Sie begründeten ihr Vorgehen mit der Vermutung, die Biopsie sei nicht repräsentativ gewesen.

Ergebnisse: Abweichend von den Mitteilungen der Autoren, die unseres Erachtens unzulässigerweise eine Anzahl von Histologien ausgeschlossen hatten, ermittelten wir aus den berichteten Rohdaten aller histologisch verifizierten Befunde Testgütekriterien an den jeweiligen Schwellenwerten, wie in Tabelle 21 dargestellt. Dabei zeigten sich geringfügig höhere Sensitivitäten beim konventionellen Pap Test gegenüber der ThinPrep-Methode. Gleichzeitig wies die ThinPrep-Methode eine insbesondere an den unteren Schwellenwerten (ASCUS+, LSIL+) höhere Spezifität auf, als der konventionelle Pap Test. Sensitivitäts- und Spezifitätsunterschiede glichen sich für beide Methoden am Schwellenwert HSIL+ auf vergleichbare Werte an.

Schlussfolgerung der Autoren: Die Autoren kamen zu der Schlussfolgerung, dass durch den Einsatz der Dünnschichtpräparation die Spezifität zur Erkennung einer Erkrankung zu erhöhen sei. Als Gründe führen sie eine Reduktion der ASCUS-Befunde und eine Absenkung falsch positiver Befunde an. In der vorliegenden Publikation waren konventionelle Pap Tests und Dünnschichtpräparate hinsichtlich ihrer Sensitivität sehr ähnlich. Wie in anderen Studien zeigte sich eine verbesserte Qualität der Objektträger durch die Dünnschichtpräparation, es waren weniger Objektträger wegen unerwünschter Beimengungen nicht auswertbar. Von den Autoren wurde jedoch auch auf einen Nachteil dieser Methode hingewiesen: durch Auswaschen von Begleitzellen (z.B. sogenannte „Tumordiathese“) die zu Informationsverlust führen, sei die Interpretation dieser Befunde erschwert.

Abschließende Bewertung: Da auch hier nur ein Teil der zytologischen Befunde histologisch verifiziert wurden, war die Ermittlung von Sensitivität und Spezifität nicht zulässig, sie wurden in der Tabelle 21 nur pro forma mitgeteilt und dürfen nicht in Betracht gezogen werden. Die Gewinnung der histologischen Proben erfolgte unter Kenntnis des zytologischen Befundes. Die vorliegende Untersuchung wurde an einem Hochrisikokollektiv durchgeführt, da nur Patientinnen mit vormals positiven Zytologien eingeschlossen wurden. Zusammenfassend müssen erhebliche Bedenken gegen interne und externe Validität der berichteten Ergebnisse der Studie ausgesprochen werden.

Tabelle 20: Evidenztabelle der Studien über Dünnschichtpräparationen von Zervikalabstrichen

Quelle	Fragestellung / Design	Technologie	Studienpopulation / Prävalenzangaben	Verwendeter Referenzstandard	Untersucher	Klassifikationssystem	Schwellenwert	Testgütekriterien	Weitere Ergebnisse	Bemerkungen	Studienqualität
Bergeron et al., 2001	Vergleich der Testgütekriterien zwischen konventionellen und dünnschichtpräparierten Zervixabstrichen RCT, Split-Sample-Design mit randomisierter Reihenfolge der Probenverarbeitung	AutoCyte PREP (AC) vs. Konventionelle Zervixabstriche	500 Patientinnen einer Kolposkopieklินิก (Hochrisikopatientinnen)	Histologischer Referenzstandard bei allen Patientinnen	Bezüglich der vorbestehenden Untersuchungsergebnisse verblindet; Auswertung durch unabhängige Untersucher	Bethesda-System für Zytologie, CIN-Klassifikation für Histologie	ASCUS+ LSIL+ HSIL+	Se (AC)= 86% Se (konv)= 82% Sp (AC)= 43% Sp (konv)= 40% PPV (AC)= 75% PPV (konv)= 78% NPV (AC)= 62% NPV (konv)= 68% Se (AC)= 70% Sp (AC)= 55% PPV (AC)= 77% NPV (AC)= 50% Se (AC)= 47% Sp (AC)= 86% PPV (AC)= 87% NPV (AC)= 45% LR+ = 1,51; LR+(konv.)= 1,35; LR- =0,32; LR-(konv)=0,46;	<u>Adequacy:</u> Unsatisfactory: 0,8% (AC) 11,6% (konv.) Limited+Unsatisfactory 7,6% (AC) 17,8% (konv.) Unterschiede statist. signifikant Bei zuerst verarbeiteter Probe, HSIL+: Se (konv)= 79%; Se (AC)=89% Sp (konv)=41%; Sp (AC)=36%; PPV (konv)=80%; PPV (AC)=71% NPV (konv)=62% NPV (AC)=65%	Spezielles Hochrisikokollektiv; Bei der Confoundingkontrolle stärkster Faktor: Reihenfolge der Probengewinnung (deutlicher Anstieg der Sensitivität der Dünnschichtpräparate, wenn zuerst verarbeitet, dann auch signifikanter Unterschied zu konventionellen Abstrichen; Deutlicher Anstieg der Spezifität bei konventionellen Präparaten, wenn zuerst verarbeitet)	B (Einschränkung durch Patientenspektrum)

Quelle	Fragestellung / Design	Technologie	Studienpopulation / Prävalenzangaben	Verwendeter Referenzstandard	Untersucher	Klassifikationssystem	Schwellenwert	Testgütekriterien	Weitere Ergebnisse	Bemerkungen	Studienqualität
Jensen et al. 2001	Vergleich der Qualität von Präparaten in Dünnschichtpräparation und konventionell. Ermittlung von benötigter Zeit zur Befundung und Sensitivität; Splitsample Studie	AutoCyte PREP™ und konventionelle Tests	1701 dänische Frauen aus kommunalem Screeningprogramm	Portiobiopsie bzw. Konisierung bei positiven Befunden, nicht beurteilbar, ob Kon-sensusurteil bei zytologischen Diagnosen zugrunde lag	k.A.	Bethesda-System	k.A.	Lt. Angaben der Autoren: AutoCytePREP™ Sensitivität: 95%. Spezifität: k.A. PPV: k.A. NPV: k.A. Konventionelle Sensitivität: 94%. Spezifität: k.A. PPV: k.A. NPV: k.A. Aus Daten der Studie FNR nicht zu ermitteln	Adequacy: AutoCytePREP: 11% SBLB: 11% Unsatisfactory: 0.1% Konventionelle SBLB: 18% Unsatisfactory: 1.8%	Keine histologische Verifikation bei negativen Befunden, mitgeteilte Sensitivität darum nicht in Betracht zu ziehen	B-C (Studie lag nur in dänisch vor, eine Einschätzung von Berichtsqualität war nur eingeschränkt möglich.)

Quelle	Fragestellung / Design	Technologie	Studienpopulation / Prävalenzangaben	Verwendeter Referenzstandard	Untersucher	Klassifikationssystem	Schwellenwert	Testgütekriterien	Weitere Ergebnisse	Bemerkungen	Studienqualität
Obweggeser et al. 2001	RCT Vergleich konventionelle und Dünnschichtpräparate bei Primärscreening (direct-to-vial)	Thin Prep™ Pap Test	N=1999 (ThinPrep™ 997, Konventionell 1002) Patientinnen aus 15 Schweizerischen privaten Gynäkologien Keine Prävalenzangaben;	Bei HSIL (n=23): Histologie von Konisations- bzw. Hysterektomie- Proben Bei LSIL / ASCUS / AGUS: Follow-up (nicht berichtet)	Auswertung der Abstriche unabhängig für jede Methode. Auswertung der Histologie bei unabhängigem Labor (Verblindung für Methode: k.A.)	Bethesda System	HSIL	Lt. Angaben der Autoren: ThinPrep™ Sensitivität 91% Spezifität: k.A. PPV: k.A. NPV: k.A. a=10, b=k.A., c=11, d=k.A. Konventionelle Sensitivität: 100% Spezifität: k.A. PPV: k.A. NPV: k.A. a=12, b=k.A., c=12, d=k.A.	Adequacy ThinPrep™ SBLB: 5.5% Unsatisfactory: 1.4% Konventionelle SBLB: 2.5% Unsatisfactory: 0% Befundraten ThinPrep™ WNL: 92.7% LSIL: 4.7% HSIL: 1.6% Ca: 0% ASCUS/AGUS+: 1.0% Konventionelle WNL: 92.9% LSIL: 3.7% HSIL: 1.8% Ca: 0.1% ASCUS/AGUS+: 1.4%	Augenmerk auf sorgfältige Abstrichentnahme, danach kein Vorteil für Dünnschichtpräparationen; Inadäquater Referenzstandard, darum Sensitivität nicht in Betracht zu ziehen; Inkomplette und teilweise inkonsistente Berichtsqualität. Keine Informationen über Erfahrungsstand der Untersucher	D (Inadäquater Referenzstandard, mangelhafte Berichtsqualität)

Quelle	Fragestellung /Design	Technologie	Studienpopulation / Prävalenzangaben	Verwendeter Referenzstandard	Untersucher	Klassifikationssystem	Schwellenwert	Testgütekriterien	Weitere Ergebnisse	Bemerkungen	Studienqualität
Park et al., 2001	Untersuchung der Effektivität des Thin-Prep™ Pap Tests im Vergleich zum konventionellen Pap Test ; Klinische Split-Sample-Studie, nicht randomisiert	Thin-Prep™ (TP) vs. Konventionelle Pap Tests	483 Patientinnen einer gynäkologischen Ambulanz des Universitätskrankenhauses Seoul;	Histologischer Referenzstandard bei 158 Patientinnen	Verblindet für Ergebnis des jeweils anderen Tests; Unabhängigkeit; k.A.	Bethesda-System für zytologische Klassifikation; CIN-Klassifikation für histologische Diagnosen	ASCUS+ LSIL+	Se (TP)= 87,5% Se (konv)= 91,6% Sp (TP)= 58,1% Sp (konv)= 46,5% PPV (TP)= 63,6% PPV (konv)= 62,3% NPV (TP)= 84,7% NPV (konv)= 76,9% Se (TP)= 87,5% Se (konv)= 90,2% Sp (TP)= 67,4% Sp (konv)= 59,3% PPV (TP)= 69,2% PPV (konv)= 65,0% NPV (TP)= 86,7% NPV (konv)= 87,9% Se (TP)= 81,9% Se (konv)= 83,3% Sp (TP)= 89,5% Sp (konv)= 86,0% PPV (TP)= 86,7% PPV (konv)= 83,3% NPV (TP)= 85,5% NPV (konv)= 86,0%		Settingansatz Hochrisiko-kollektiv beachten, damit PPV und NPV als prävalenzabhängige Maße nicht auf Screeningpopulationen übertragbar; Sensitivität und Spezifität wurden nur pro Forma berichtet, dürfen bei inadäquatem Referenzstandard jedoch nicht in Betracht gezogen werden; Erhebliche Zweifel an interner und externer Validität der Untersuchungsergebnisse.	D (Unzureichende methodische Qualität bei inadäquatem Referenzstandard, anzunehmende erhebliche Verzerrung von Ergebnissen)

I C 5. 3. 2 Nicht berücksichtigte Publikationen

Folgende Publikationen wurden zunächst als potentiell relevant identifiziert, bei einer näheren Betrachtung wurden sie ausgeschlossen. Sie werden in der folgenden Tabelle (Tabelle 21) mit Nennung der wichtigsten Ausschlussgründe zusammengefasst.

Tabelle 21: *Dünnschichtpräparationstechniken im Zervixkarzinom-Screening: Ausgeschlossene Studien*

Studie	Ausschlussgrund
Boman et al.: Pourquoi nous préférons la technique en couche mince aux frottis cervico-utérins conventionnelles. <i>Clin Exp Path</i> 1999;47:81-87	Bei der Studie handelt es sich um eine Machbarkeitsstudie, die keine Outcomeparameter unseres Interesses berichtete (Adequacy).
Marino JF, and Fremont-Smith M: Direct-to-Vial Experience with AutoCyte PREP in a Small New England Regional Cytology Practice. <i>The Journal of Reproductive Medicine</i> 2001; 46: 353-358.	Bei der Studie handelt es sich um eine Machbarkeitsstudie, die keine Outcomeparameter unseres Interesses berichtete (Befundraten, Adequacy).
Massarani-Wafai et al.: Evaluation of Cellular Residue in the ThinPrep™ PreservCyt™ Vial. <i>Diagnostic Cytopathology</i> 2000; 23; No. 3: 208-212.	Die Studie untersuchte den Benefit, den ein zusätzlich im Dünnschichtverfahren präparierter Objektträger für die diagnostische Aussagekraft dieser flüssigkeitsbasierten Methode liefert. Dieser Outcomeparameter entsprach nicht unseren Einschlusskriterien.
Monsonogo J, Autillo-Touati A, Bergeron C, Dachez R, Liaras J, Saurel J, Zerat L, Chatelain P, and Mottot C : Liquid-based cytology for primary cervical cancer screening: a multi-centre study. <i>British Journal of Cancer</i> 2001; 84(3): 360-366.	Bei der Studie handelt es sich um eine Machbarkeitsstudie, die keine Outcomeparameter unseres Interesses berichtete (Befundraten, Adequacy).
Shield et al.: Improving Cervical cytology screening in a remote, high risk population. <i>MJA</i> 1999; 170: 255-258.	Die Studie untersuchte den Einfluss des Einsatzes von Dünnschichtpräparationen mit ThinPrep™ auf die Senkung der Anzahl „unsatisfactory“ Abstriche in entlegenen Kommunen in Queensland, Australien. Diese high-risk-Population weist Spezifika auf, die durch einen erhöhten Anteil entzündlicher und maligner Veränderungen nicht auf eine deutsche Screeningpopulation übertragbar sind. Darüber hinaus machten methodische Mängel das Auftreten systematischer Fehler wahrscheinlich.

I C 5. 4 Zusammenfassung der Ergebnisse – Effektivität

In diesem Abschnitt wird eine Kurzzusammenfassung der Ergebnisse der berücksichtigten Publikationen in Form von Übersichtstabellen vorgenommen.

Die hier dargestellten Ergebnisse werden im anschließenden Kapitel A.6. Diskussion kritisch gewürdigt.

I C 5. 4. 1 HTA-Berichte, systematische Reviews und Metaanalysen

In der folgenden Tabelle (Tabelle 22) werden die wichtigsten Ergebnisse der systematischen Technologiebewertung aus sekundären Publikationen zusammengefasst.

Übereinstimmend wurde in allen HTA-Berichten eine unzureichende Evidenzlage für Testgütekriterien aller neuen Screeningtechnologien, d.h. sowohl für Dünnschichtpräparationen, als auch für computergestützte Auswertung von Pap Tests, sowie die überwiegend mangelhafte Qualität bei den durchgeführten klinischen Studien konstatiert.

Für Dünnschichtpräparationen liegen Hinweise vor, dass diese Verarbeitungstechnik einen höheren Anteil adäquater Abstrichproben liefern kann.

I C 5. 4. 2 Primärstudien

Tabelle 23 enthält einen zusammenfassenden Überblick über die Ergebnisse aus den Primärstudien.

Tabelle 22: Zusammenfassung der berücksichtigten Sekundärpublikationen

Quelle	Ergebnisse	Kritik der Autoren	Schlussfolgerungen der Autoren
HTA-Berichte mit Bewertung aller neuen Technologien (Dünnschichtpräparationstechniken und computergestützte Auswertung von Pap Tests)			
AHTAC (Australien), 1998	<ul style="list-style-type: none"> - Keine Evidenz für Überlegenheit von Dünnschichtpräparationen oder computergestützter Auswertung von Pap Tests; 	<ul style="list-style-type: none"> - Fehlende Verifikation der zytologischen Testergebnisse; - Zu wenige Studien; - Methodische Mängel bei Studien. 	<ul style="list-style-type: none"> - Keine Senkung von Inzidenz und Mortalität durch neue Technologien erwartet.
Broadstock, 2000; NZHTA (Neuseeland)	<ul style="list-style-type: none"> - Keine Evidenz für Überlegenheit von Dünnschichtpräparationen oder computergestützter Auswertung von Pap Tests; <p>(Insbesondere keine Evidenz für Verbesserung der Sensitivität, keine Aussagen über Spezifität möglich bei beiden Technologien).</p>	<ul style="list-style-type: none"> - Keine RCT's mit Outcome invasives Karzinom; - Studien methodisch mangelhaft (Population, Design, Referenzstandards) 	<ul style="list-style-type: none"> - Einführung neuer Technologien in Neuseeland nicht befürwortet; - Forderung nach neuen Studien zur Klärung der Evidenzlage.
Mc Crory et al., 1999; AHCPR (U.S.A.)	<ul style="list-style-type: none"> - Sensitivität des <i>konventionellen Pap Tests</i> mit 51% schlecht; - <i>Dünnschichtpräparationen</i>: bei kolposkopisch-histologischer Verifikation vergleichbar mit Pap Test, bei zytologischem Referenzstandard Sensitivität gegenüber konventionellem Pap Test verbessert, keine Aussagen zur Spezifität. - <i>Computergestützte Auswertung</i>: bei kolposkopisch-histologischer Verifikation vergleichbar mit Pap Test, bei zytologischem Referenzstandard Sensitivität gegenüber konventionellem Pap Test verbessert, keine Aussagen zur Spezifität. 	<ul style="list-style-type: none"> - Starke Variabilität der Studienqualität, meiste - Studien methodisch mangelhaft, insbesondere mangelhafte Referenzstandards. 	<ul style="list-style-type: none"> - Unsichere Evidenzlage für Steigerung der Effektivität durch neue Technologien bei höheren Kosten; - Weitere Studien unter Einsatz von akzeptablen Referenzstandards erforderlich.

Quelle	Ergebnisse	Kritik der Autoren	Schlussfolgerungen der Autoren
HTA-Bericht mit Bewertung der Dünnschichtpräparationstechniken			
Payne et al., 2000 ; NCCHTA (Großbritannien)	<ul style="list-style-type: none"> - Weniger inadäquate oder eingeschränkt beurteilbare Proben bei Dünnschichtpräparaten; - Unsichere Evidenzlage zu Parametern Sensitivität und Spezifität. 	<ul style="list-style-type: none"> - Mangelhafte Referenzstandards; - Zu wenige Studien an Screeningpopulationen; - Tendenz zum Diagnose-Upgrading bei Dünnschichtpräparaten. 	<ul style="list-style-type: none"> - Weitere Studien in RCT-Qualität mit histologischen Referenzstandards zur Bestimmung von Sensitivität und Spezifität erforderlich, - Einsatz der Dünnschichtpräparationen lässt Abnahme der inadäquaten Abstriche erwarten.
Systematischer Review und Metaanalyse über ThinPrep™-Dünnschichtpräparationstechnik			
Metaanalyse von Bernstein et al., 2001	<ul style="list-style-type: none"> - Verbesserung der Adequacy durch ThinPrep™; - Detektionsraten bei ThinPrep™ teilweise besser 	<ul style="list-style-type: none"> - Heterogenität der Studien bezüglich Populationen, - Fehlende histologische Verifikation zur Bestimmung der Testgütekriterien. 	<ul style="list-style-type: none"> - ThinPrep™ lieferte adäquatere Proben, keine Aussagen über Sensitivität und Spezifität möglich.

Tabelle 23: Zusammenfassung der berücksichtigten Primärpublikationen

Quelle	Technologie	Ergebnisse
- Primärpublikationen über Dünnschichtpräparationen		
Bergeron et al. 2001	AutoCyte PREP™	<ul style="list-style-type: none"> - Geringe Verbesserung der Sensitivität durch AutoCyte PREP™ nicht statistisch signifikant. - Bei zuerst verarbeiteten Proben Sensitivität für AutoCyte PREP™ etwas besser als Pap Test (statist. sign.); Spezifität schlechter als bei Pap Test; - Adequacy besser bei AutoCyte PREP™ als bei Pap Test
Jensen et al. 2001	AutoCyte PREP™	<ul style="list-style-type: none"> - Verbesserung der Adequacy, - Keine nennenswerte Verbesserung der Sensitivität von AutoCyte PREP™
Obwegeser et al. 2001	ThinPrep™	<ul style="list-style-type: none"> - Sensitivität von ThinPrep™ schlechter als Pap Test; - Adequacy bei ThinPrep™ schlechter als Pap Test
Park et al. 2001	ThinPrep™	<ul style="list-style-type: none"> - Sensitivität bei ThinPrep™ etwas schlechter als Pap Test, - Spezifität bei ThinPrep™ geringfügig besser als Pap Test
- Primärpublikationen über computergestützte Auswertung von Pap Tests mit AutoPap™		
Alasio et al. 2001	AutoPap™ als Primärscreener	<ul style="list-style-type: none"> - Sensitivität AutoPap™ 97%, - Spezifität AutoPap 40%

Quelle	Technologie	Ergebnisse
Renshaw et al. 2000	AutoPap™ als Rescreener	- AutoPap™-Rescreening senkt Rate Falsch Negativer stärker als 10%-iges randomisiertes manuelles Rescreening

Bei den berücksichtigten Primärpublikationen erhielt nur eine Studie den Qualitätsscore B (Bergeron et al. 2001), eine Studie lag nur in der Originalsprache Dänisch vor (Jensen et al. 2001) und der Qualitätsscore konnte darum nur mit Einschränkungen bestimmt werden (B-C), alle anderen erhielten aufgrund inadäquater Referenzstandards und/oder bei schwerwiegenden methodischen Mängeln den Qualitätsscore D.

Die hier bewerteten Primärstudien über Dünnschichtpräparationen haben die Evidenzlage, wie sie in den HTA-Berichten beschrieben wurde, nicht verändert. Insbesondere ergaben sie keinen Erkenntniszuwachs hinsichtlich der Testgütekriterien.

Auch bei der Bewertung des AutoPap™-Verfahrens, der computergestützten Auswertung von Pap Tests, konnte durch die beiden identifizierten Studien keine klinische Überlegenheit dieser Technologie gegenüber konventionellen Pap Tests demonstriert werden, da die Studien methodischen Mängeln unterworfen waren, die eine kritische Einschätzung der berichteten Testgütekriterien erforderlich machte.

I C 6 Diskussion

Der vorliegende Bericht unterliegt der Zielsetzung, neue Technologien des Zervixkarzinom-Screenings auf deren medizinische Effektivität zu prüfen.

Ausgehend von einer systematischen Literaturrecherche, deren Ergebnisse nach Anwendung der Ein- und Ausschlusskriterien im Kapitel I A.5. im Einzelnen berichtet wurden, werden die aus internationaler Forschung stammenden Ergebnisse im folgenden Kapitel diskutiert. Dabei folgt der Aufbau des Kapitels dem des Ergebnisteils mit Bezugsrahmen, Methodik und inhaltlichen Aspekten.

I C 6. 1 Bezugsrahmen

In zunehmendem Maße werden Ärzte sowie die breite Öffentlichkeit direkt von den Herstellern neuer Screeningtechnologien, deren Kosten nicht von gesetzlichen Krankenversicherungen übernommen werden, beworben. Die neuen Technologien im Bereich der Vorsorgeuntersuchungen gegen Gebärmutterhalskrebs befinden sich, betrachtet man die gegenwärtige Phase im Technologiezyklus, in Deutschland in der Einführungsphase, einer Phase, die naturgemäß mit offensiven Marketingstrategien vergesellschaftet ist (vgl. Kapitel I A.2.5.). Betrachtet man das Gesundheitswesen unter rein marktwirtschaftlichen Gesichtspunkten, die freie Interaktionen zwischen Anbietern und Konsumenten vorsehen, erscheinen neue Technologieangebote an frei entscheidende Konsumenten zunächst als legitim. Man übersieht dabei jedoch den Professionalitätsaspekt der Medizin (Lee & Brennan; 2002). Die Medizin steht in einer Verantwortung, die auf der Grundlage altruistischer Empfehlungen von Ärzten an ihre individuellen Patienten basiert. Das bedeutet, dass die Empfehlungen für oder gegen die Anwendung einer neuen Technologie über eine bloße Aufklärung über mögliche nachteilige Folgen hinausgehen. Denn sowohl die falsch negativen Befunde, die bei Untersuchten ein falsches Gefühl der Sicherheit auslösen, als auch falsch positive Ergebnisse, die gesunde Menschen einem Risiko aussetzen, das aus teilweise invasiven Folgeuntersuchungen erwächst, müssen bei der Bewertung neuer Technologien im Vergleich zu herkömmlichen Verfahren in Betracht gezogen werden. Das Angebot unsicherer Testverfahren steht somit im Gegensatz zur Ethik ärztlichen Handelns (Lee & Brennan; 2002).

Von den Kostenträgern sind gemäß den Vorgaben des Gesetzgebers zur Berücksichtigung des medizinischen Fortschritts unter gleichzeitiger Beachtung des Wirtschaftlichkeitsgebots (§2, Abs.1, SGB V) bei zunehmender Verschärfung der finan-

ziellen Ressourcen ebenfalls rationale, evidenzbasierte Entscheidungen gefordert. Sie unterliegen gleichzeitig einem wachsenden öffentlichen Druck zur Übernahme der Kosten neuer diagnostischer und therapeutischer Verfahren, der nicht nur auf erfolgreiches Industriemarketing zurückzuführen ist, sondern auch von Aussagen aus Print-, Fernseh- und Internetmedien unterstützt wird. So wurde z.B. in einer Sendung des „ARD-Ratgeber Gesundheit“ davon berichtet, dass eine Kombination von Computer und Arzt die Treffsicherheit der Vorsorge gegen Gebärmutterhalskrebs von 85% auf 99% anheben soll. Diese Methode würde von den Krankenkassen nicht bezahlt, weil sie mit 95 DM wesentlich teurer sei, als die herkömmliche (ARD-Ratgeber Gesundheit; 1997). Oder es werden Expertenmeinungen präsentiert, deren Haltung gegenüber dem Pap Test im WDR-Fernsehen so formuliert wurde: „Der Pap Test ist zwar nicht schlecht...Aber in 70% der Fälle übersieht er Krebsstufen. Deshalb sollten empfindlichere Tests zur Vorsorge zugelassen werden...“ (WDR-Fernsehen; 2001).

Neben dem möglichen Einfluss dieser und ähnlicher Meldungen auf die öffentliche Meinungsbildung mit Auswirkungen auf politische Entscheidungsträger stellen Fehlinformationen, wie die oben zitierten, eine Gefahr für die Effektivität des gesamten Vorsorgeprogramms gegen Gebärmutterhalskrebs dar. Auch Informationen, die einseitig die Limitationen des Screeningtests in den Vordergrund stellen, führen zur Abnahme der Teilnahmeraten, die eine wesentliche Determinante eines erfolgreichen Screeningprogramms bilden (Jepson et al.; 2000). Ein großer Teil von Frauen, die in ihrem Leben ein invasives Zervixkarzinom entwickeln, haben niemals oder nicht in den erforderlichen Zeitabständen am Screeningprogramm teilgenommen (McCroly et al.; 1999). Aus dieser Tatsache ist ersichtlich, dass neben der hier vorgelegten Technologiebewertung für die Effektivität eines einzelnen Screeningtests, Aspekte der Screeningstrategie einen erheblichen Einfluss auf die Effektivität des gesamten Screeningprogramms nehmen.

Einen weiteren wesentlichen Einflussfaktor der Effektivität stellen innere Qualitätsfaktoren bei der Durchführung im klinischen Routinebetrieb, d.h. unter Alltagsbedingungen im Gegensatz zu den hier ausgewerteten Publikationen unter Studienbedingungen dar. Unter den folgenden Punkten sind einige dieser Aspekte in kurzer Form ausgeführt:

- Während in den U.S.A. eine Obergrenze von maximal 85 Präparaten pro Zytologieassistentin pro 24 Stunden gesetzlich vorgeschrieben ist (Schneider; 1995), existiert hierzu in Deutschland nur eine Empfehlung der Bundesärztekammer über maximal 10 Objektträger pro Stunde (Leitlinie der BÄK zur Qualitätssicherung zytologischer Untersuchungen...“ vom 17.12.1993). Die Einhaltung dieser Empfehlung wird für einige Laboratorien angezweifelt (Schenck; 2000) und es wird befürchtet, dass bei überhöhter Zahl von durchgemusterten Objektträgern

diese nicht vollständig gescreent werden oder dass Befunde infolge Übermüdung übersehen werden.

- Da Erfahrungs- und Ausbildungsstand von Zytologischen Assistenten eine zentrale Bedeutung bei der Beurteilung von Präparaten haben, können auch Änderungen der Ausbildungsrichtlinien bei dieser Berufsgruppe Einfluss auf die Effektivität des Screeningprogramms ausüben. Mit der Rahmenvereinbarung über die Ausbildung und Prüfung zum technischen Assistenten/ zur technischen Assistentin an Berufsfachschulen (Beschluss der Kultusministerkonferenz vom 12.6.1992 in der Fassung vom 26.6.1998) wurde die zuvor eigenständige Berufsausbildung zum Zytoassistenten in die Berufsausbildung der MTA's integriert. Das hatte eine deutliche Reduktion der Lehrinhalte gynäkologischer Zytologie und die Schließung von Zytoassistentenschulen zur Folge. Vom Verband Deutscher Cytologisch Tätiger Assistenten e.V. wird als Folge nicht nur mit Qualitätseinbußen bei den Auszubildenden gerechnet, sondern darüber hinaus mit einem erheblichen Mangel an fertig ausgebildeten Zytoassistenten (persönliche Kommunikation E. Müller-Leibenger, Vors. d. VDCA, vom 20.02.2002). Ein derartiger Mangel an Zytoassistenten hat in Japan bereits vor Jahren zur Einführung computergestützter Auswertungssysteme beigetragen (AHTAC-Report 1998). Zur erwarteten Marktentwicklung von AutoPap™ befragt, teilte uns Medite® mit:

„Vor dem Hintergrund, dass in Deutschland fast alle Schulen, die sich mit der Ausbildung von Cytologisch-Technischen-AssistentInnen befassen, geschlossen wurden, ist bereits jetzt schon eine Verknappung von qualifiziertem Personal festzustellen. Vor diesem Hintergrund erwarten wir eine positive Marktentwicklung ...“ (Angabe Fa. Medite GmbH, Generalvertrieb Deutschland für TriPath Imaging; pers. Kommunikation H. Goldbecker, 15.01.2002).

I C 6. 2 Diskussion der Methodik

I C 6. 2. 1 Literaturrecherche

Das Problem einer systematischen Literaturrecherche für diagnostische Studien ist bereits von mehreren Autoren thematisiert worden. So weist Deeks (2001) darauf hin, dass eine Suche in elektronischen Datenbanken weit weniger produktiv ist für diagnostische Studien als für randomisierte klinische Therapiestudien wegen einer inkonsistenten Verschlagwortung der Publikationen. Um möglichst keine relevanten Studien zu übersehen, wurde eine Kombination von MeSH- und Freitextbegriffen in verschiedenen Schreibweisen gewählt. Dabei wurden zahlreiche irrelevante Publikationen identifiziert, die in einem zweiten manuellen Selektionsschritt ausgesondert

werden mussten. Diese Suche wurde durch eine Handsuche ergänzt. Es ist wenig wahrscheinlich, dass von uns wesentliche publizierte Studien zur Evaluation der neuen Screeningtechnologien übersehen wurden.

Offen bleibt jedoch die Problematik des sogenannten *Publication Bias*, der dadurch entsteht, dass vermehrt Studien in wissenschaftlichen Journalen publiziert werden, die ermutigende Resultate zeigen (Lijmer et al.; 1999). Da es keine zentrale Registrierung für die Durchführung von diagnostischen Studien gibt, kann nicht abgeschätzt werden, wie groß der Anteil nicht publizierter Studien ist.

I C 6. 2. 2 Allgemeine methodische Probleme in der Bewertung der Qualität von Primärstudien

Wie bereits in Kapitel I A.4.2 dargelegt, existiert kein allgemein anerkanntes Evaluationskonzept für Primärstudien diagnostischer Tests. Von verschiedenen Arbeitsgruppen sind Vorschläge erarbeitet worden, die eine phasenweise Evaluation diagnostischer Tests (Flynn & Adams; 1996, Sackett & Haynes; 2002) in Anlehnung an die in Therapiestudien übliche 4-Phasen-Einteilung vorschlagen. Diese Diskussion ist noch nicht abgeschlossen. Gemäß dem Konzept von Sackett & Haynes (2002) waren die hier ausgewerteten Studien den Phasen I und II zuzuordnen (siehe Tabelle 11).

Auch für die Bewertung der Qualität von Studien diagnostischer Tests, bei der es um die Beurteilung interner und externer Validität der Studienergebnisse geht, gibt es keine allgemeingültigen Standards. In der vorliegenden Arbeit wurde auf einen Bewertungsmaßstab zurückgegriffen, der bereits in anderen HTA-Berichten eingesetzt wurde und der auf einem Vorschlag von Flynn & Adams (1996) beruht. Da es bislang nur wenig empirische Evidenz mit quantifizierenden Aussagen über ergebnisverzerrende Effekte (wie z.B. bei Lijmer et al.; 1999) gibt, ist bewusst auf den Einsatz von Punktescores und Schwellenwerten im Rahmen von Checklisten als Einschlusskriterium verzichtet worden.

Im folgenden Abschnitt werden methodische Probleme angesprochen, denen eine Vielzahl von ausgewerteten Studien unabhängig von der untersuchten Technologie (Dünnschichtpräparationen oder computergestützte Auswertung von Pap Tests) unterlagen, die z.T. bereits in vorausgegangenen HTA-Berichten internationaler Organisationen identifiziert wurden und deren Beachtung bei der Durchführung weiterer klinischer Studien gefordert wurde.

Studienpopulationen

Die Zusammensetzung der Studienpopulation hinsichtlich der Krankheitsprävalenz stellt einen wichtigen Parameter für die Übertragbarkeit von Ergebnissen dar. Das betrifft z.B. primär prävalenzabhängige Outcomeparameter wie positive Vorhersagewahrscheinlichkeiten (PPV) und negative Vorhersagewahrscheinlichkeiten (NPV). Mit steigender Prävalenz steigen auch die im Test ermittelten PPV's (Gordis; 2001). Das bedeutet, dass PPV's, die an Hochrisiko-Studienpopulationen ermittelt wurden, nicht auf Screeningpopulationen übertragbar sind. Alle von uns ermittelten Primärstudien, die diese Outcomeparameter PPV/NPV berichteten oder bei denen sie aus berichteten Rohdaten berechenbar waren, arbeiteten an Hochrisikokollektiven (Bergeron et al.; 2001, Park et al.; 2001, Alasio et al.; 2001: hier wurde eine Screeningpopulation mit Hochrisikopatientinnen „angereichert“). Die berichteten PPV-/NPV-Werte sind somit auf deutsche Screeningpopulationen nicht übertragbar.

Auch die Testgütekriterien Sensitivität und Spezifität, die nicht prävalenzabhängig sind, werden durch ergebnisverzerrende Effekte, also Formen von systematischen Fehlern, deren Ausmaß mit der Prävalenz zusammenhängt, beeinflusst (McCorry et al.; 1999). Diese Autoren fanden in ihrer Metaanalyse konventioneller Pap Tests eine deutliche Ergebnisverzerrung von Sensitivität und Spezifität durch einen *Work-up-Bias* (Systematischer Fehler, der zu Verzerrungen von Ergebnissen führt, bei dem insbesondere in Screeninguntersuchungen bei positiven Testergebnissen eine verstärkte Nachuntersuchung der Betroffenen erfolgt. Diese werden einer Referenzdiagnostik, wie z.B. einer histologischen Diagnose zugeführt, während eine histologische Nachuntersuchung von negativen Screeningbefunden unterbleibt.), der entsteht, wenn die Anwendung des Referenzstandards in Abhängigkeit von den Ergebnissen des Screeningtests erfolgt. Dabei konnten sie zeigen, dass bei vorliegendem Work-up-Bias die Sensitivität eher überschätzt und die Spezifität eher unterschätzt werden. Da der Work-up-Bias mit steigender Prävalenz zunimmt, wird die Ergebnisverzerrung in hochprävalenten (Hochrisiko-) Populationen stärker ausfallen, als in Screeningpopulationen niedriger Prävalenz (McCorry et al.; 1999).

Auch Knottnerus et al. (2002) diskutierten den Einfluss der Zusammensetzung der Studienpopulation. Sie wiesen darauf hin, dass die Sensitivität eines Tests überschätzt wird, wenn sie an einer Studienpopulation erhoben wird, die mehr und schwerer erkrankte Individuen aufweist, als die Population, an welcher der Test eingesetzt werden soll – in unserem Falle also die Überschätzung der Sensitivität durch Studienpopulationen aus Hochrisikokollektiven bei Alasio et al. (2001), Bergeron et al. (2001) und Park et al. (2001). Knottnerus et al. (2002) bezeichneten derartige Verzerrungen als *Spectrum Bias*. Häufig treten in Verbindung mit dem *Spectrum Bias* auch Selektionsverzerrungen auf, die dadurch entstehen, wenn Individuen in Abhängigkeit von einem vorausgegangenen Testergebnis in eine Studie aufgenom-

men werden (*Selection Bias*) (Knottnerus et al.; 2002), hier in den Untersuchungen von Bergeron et al. (2001) und Park et al. (2001).

Auch bei Irwig et al. (2002) wird der Einfluss des Settings auf die Sensitivität in gleicher Weise diskutiert. Sie vertreten die Ansicht, dass die Übertragbarkeit von z.B. Sensitivitätsangaben von einem Setting auf ein anderes an einige wesentliche Voraussetzungen geknüpft ist und hoben hervor, dass die Verteilung der Ergebnisse eines Tests bei Erkrankten in Durchschnitt wie auch Verteilungsmuster konstant sein muss. Diese Forderung wird nicht erfüllt, wenn es zu einem Wechsel im Erkrankungsspektrum kommt, in unserem Falle also die Sensitivität bei einem Hochrisikokollektiv ermittelt wird und auf eine Screeningpopulation angewendet werden soll.

Als weiteres Problem bei der Auswertung der Primärstudien erwies sich die meist fehlende Prävalenzangabe in der Publikation, was nicht nur einen Mangel an Berichtsqualität darstellt, sondern darüber hinaus eine Einschätzung ergebnisverzerrender Effekte erschwert. In keiner der ausgewerteten Studien wurden ausreichende Angaben zur Art der Rekrutierung der Studienteilnehmer gemacht, insbesondere nicht dazu, ob die teilnehmenden Frauen sich selbst für die eine oder die andere Methode entschieden oder der sie behandelnde Arzt diese Entscheidung traf. Broadstock (2000) diskutiert dazu: Bei Frauen, die sich selbst für einen Test entscheiden, der zusätzliche finanzielle Aufwendungen bedeutet, könne nicht davon ausgegangen werden, dass sie derselben Population entstammen wie Frauen, die sich für den (kostenfreien) konventionellen Test entscheiden. Auch in Studien, bei denen die Zuweisung durch Auswahl des behandelnden Arztes getroffen wurde, konnten Auswahlverzerrungen mit nachfolgender Ergebnisverzerrung nicht ausgeschlossen werden. (Broadstock; 2000).

Studiendesigns

Unter den eingeschlossenen Studien befanden sich nur zwei randomisierte kontrollierte klinische Studien (RCT's) (Bergeron et al.; 2001, Obwegeser et al.; 2001), alle anderen waren nicht randomisierte klinische Studien. Dabei ist in der Studie von Bergeron nur für die Reihenfolge der Probenverarbeitung randomisiert worden, streng genommen ist also auch diese Studie kein RCT.

Ein besonderes Problem bei der Auswahl des Studiendesigns taucht bei der Untersuchung von Dünnschichtpräparationen auf. Hier finden zwei prinzipiell unterschiedliche Designs Anwendung: das Split-Sample-Design (SSD) und das Direct-to-Vial-Design (DVD). Beim SSD wird zunächst ein konventioneller Pap Abstrich gefertigt, anschließend wird das Abstrichentnahmeanstrument mit verbliebenem Abstrichmaterial in einer Phiole mit Fixierflüssigkeit ausgeschwenkt. Beim DVD wird das Entnah-

meinstrument direkt in die Phiole mit Fixierflüssigkeit gegeben, ohne vorher einen konventionellen Ausstrich angefertigt zu haben. Der Vorteil vom SSD besteht in einem intraindividuellen Vergleich, da beide Tests an derselben Frau vorgenommen werden. Das bedeutet, dass sichergestellt ist, dass beide „Studienpopulationen“ derselben Grundgesamtheit entstammen. Der Nachteil besteht darin, dass die zuletzt durchgeführten Dünnschichtpräparationen nicht mehr genügend repräsentatives Zellmaterial enthalten und damit durch dieses Studiendesign benachteiligt werden. Dazu lieferte die hier präsentierte Studie von Bergeron et al. (2001) einen interessanten Ansatz. Unter Einsatz des SSD wurde durch eine Randomisierung die Reihenfolge der Probenverarbeitung festgelegt, d.h. es wurde per Zufallsauswahl bestimmt, ob der Abstrich zuerst in konventioneller Weise ausgestrichen wurde oder ob er zuerst als Dünnschichtpräparation aufgearbeitet wurde. Wurden die Präparate zuerst in Dünnschichtpräparationstechnik aufgearbeitet, lag ihre Sensitivität höher als die in zweiter Linie verarbeiteten Pap Tests. Wurden zunächst Pap Tests gefertigt, lag deren Spezifität höher als bei den in zweiter Linie angefertigten Dünnschichtpräparaten. Weitere Untersuchungen müssten jedoch zeigen, ob konventionelle Pap Tests, die in zweiter Linie hergestellt werden, nachdem das Abstrichinstrument in einer Fixierflüssigkeit ausgeschwenkt wurde, nicht anderen Limitationen unterliegen und somit neue Verzerrungen der Ergebnisse in Betracht gezogen werden müssen.

Referenzstandard

Als wichtigstes methodisches Problem erwies sich in fast allen von uns ausgewerteten Studien der fehlende oder zumindest inkomplette Referenzstandard. Auch auf dieses Problem war schon von allen Autoren der anderen HTA-Organisationen, deren Berichte wir in unsere Auswertung aufnahmen, mit Nachdruck hingewiesen worden.

Um Sensitivität und Spezifität eines neuartigen Testverfahrens bestimmen zu können, ist es erforderlich, den Wahrheitsgehalt der ermittelten Diagnosen mit einer Methode, die prinzipiell nicht den gleichen Fehlerquellen unterworfen ist, zu prüfen. Die bloße Übereinstimmung von Diagnosen von zwei zytologischen Testverfahren birgt die Gefahr, dass konkordant positive/negative Befunde darum „nicht richtig“ sein müssen, weil sie prinzipiell auch beide falsch sein können. Als von der Zytologie unabhängige Methode und „Goldstandard“ der Verifikation gilt die Histologie (Broadstock; 2000, McCrory et al.; 1999). Darum wäre eine komplette histologische Kontrolle aller Befunde, also auch der negativen Befunde zu fordern, um Sensitivität und Spezifität zu ermitteln. Frauen mit negativen Testergebnissen einer invasiven Untersuchung mit prinzipiellen Komplikationsmöglichkeiten zu unterziehen, ist jedoch aus ethischen Gründen nicht aufrecht zu halten. Darüber hinaus unterliegt auch die Histologie potentiellen Fehlerquellen. So kann z.B. nicht gewährleistet werden, dass immer eine repräsentative Gewebeprobe gewonnen werden kann, d.h. dass das

Präparat auch tatsächlich den interessierenden Bereich enthält. Aus diesen Gründen sind von der Intersociety Working Group for Cytology Technologies (ISWG (Intersociety Working Group for Cytology Technologies: Ausschuss, der aus führenden Mitgliedern der folgenden sechs Gesellschaften konzipiert wurde: American Society for Cytotechnology, American Society of Clinical Pathologists, American Society of Cytopathology, College of American Pathologists, International Academy of Cytology, Papanicolaou Society of Cytopathology.)) methodische Leitlinien zur Konzeption von klinischen Studien zur Untersuchung von Testmethoden in der gynäkologischen Zytologie in Form eines Konsensuspapiers verabschiedet worden. Darin wird ein zytologischer Referenzstandard empfohlen, der von einem unabhängigen Expertengremium im Konsensusprozess erhoben wird und der die histologische Verifikation eines signifikanten Teils der hochgradigen Dysplasien (sic!) einschließt (Intersociety Working Group for Cytology Technologies 1997 und 1998). Von McCrory et al. (1999) wurde dieser Referenzstandard als Surrogat akzeptiert, sie schränkten jedoch die zu ermittelnden Outcomeparameter ein. So können nach Chock et al. (1997, in McCrory et al., 1999) zwar keine Sensitivität und Spezifität ohne komplette histologische Verifikation ermittelt werden. Zulässig ist jedoch, relative TPR und FPR zu berechnen, sofern beide Tests unabhängig voneinander appliziert worden seien. Diese Bedingung der Unabhängigkeit ist jedoch z.B. bei computerisiertem Rescreening von vormals als negativ befundenen Objektträgern nicht erfüllt (McCrory et al.; 1999).

Unter den identifizierten Primärstudien war eine Studie mit kompletter histologischer Verifikation der Diagnosen: Bergeron et al. (2001). In der Split-Sample Studie konnten keine statistisch signifikanten Unterschiede von Sensitivität und Spezifität für die mit konventionellen Pap Abstrichen verglichenen AutoCyte PREP™-Präparate ermittelt werden. Erst als für die Reihenfolge der verarbeiteten Präparate getrennte Auswertungen vorgenommen wurden, ergaben sich Vorteile der Sensitivität für AutoCyte PREP™ bei gleichzeitig verminderter Spezifität. Wie bereits oben angesprochen, ist dieses Design, einen konventionellen Pap Abstrich in zweiter Linie herzustellen, nicht ausreichend erprobt, um eine etwaige Benachteiligung der Pap Abstriche auszuschließen. So liegen z.B. keine validen Aussagen vor, ob bei dieser Reihenfolge ausreichend repräsentatives Material bei den in zweiter Linie hergestellten Pap Abstrichen vorgelegen hat.

In keiner weiteren Studie wurde ein adäquater Referenzstandard eingesetzt, um andere Testgütekriterien, wie etwa FNR oder FPR zu ermitteln. Auch die bei Renshaw et al. (2000) berichteten FNR sind bei kritischer Würdigung nicht akzeptabel. Die Autoren setzten zwar ein zytologisches Konsensusverfahren ein, nahmen jedoch nicht die geforderten histologischen Verifikationen bei einem signifikanten Anteil hochgradig positiver Befunde vor.

Darüber hinaus wurden bei Alasio et al. (2001), Jensen et al. (2001) und Obwegeser et al. (2001) keine Angaben zur Verblindung der Untersucher des Referenztests gegenüber dem Ergebnis der zytologischen Untersuchung gemacht. Lijmer et al. (1999) fanden in ihrer Analyse von 218 diagnostischen Studien nach ergebnisverzerrenden Einflüssen der Methodik, dass Studien, deren Untersucher unter Kenntnis der vorangegangenen Testergebnisse die Verifikation durch den Referenzstandard vornahmen und Studien, die hierzu keine Angaben machten, die Leistungsfähigkeit von diagnostischen Tests überschätzten. Ebenso berichteten sie über eine Überschätzung der Testgütekriterien bei differierendem Einsatz des Referenzstandards für positive und negative Befunde.

Testschwellenwerte

Bereits bei Broadstock (2000) wurde explizit auf das Problem der Festlegung von Schwellenwerten, ab welchem Schweregrad von Veränderungen ein zytologischer Befund als positiv zu bezeichnen ist, hingewiesen. Hintergrund dieser Überlegungen ist die Tatsache, dass ein großer Anteil niedriggradiger Veränderungen zu Spontanremissionen neigt, d.h. keinen Krankheitswert im eigentlichen Sinne besitzt. McCrory et al. (1999) wiesen über diese expliziten Schwellenwerte hinaus auf das Problem impliziter Schwellenwerte hin, d.h. die subjektive Einschätzung der UntersucherInnen, ab wann sie einen Befund als pathologisch einschätzen. In der einzigen Primärstudie, die aufgrund eines adäquaten Referenzstandards erhobene Testgütekriterien berichtete (Bergeron et al. 2001), fielen, wie erwartet, die Sensitivitätswerte für AutoCyte PREP™ mit steigender Testschwelle bei gleichzeitiger Zunahme der Spezifität für alle Proben, unabhängig von der Reihenfolge der Verarbeitung (Sensitivität bei ASCUS+= 89%, bei LSIL+= 69%, bei HSIL+= 47%; Spezifität bei ASCUS+= 43%, bei LSIL+= 60%, bei HSIL+= 86%).

Während bei Broadstock (2000) ein Schwellenwert von HSIL+ gefordert wurde, da diese Testschwelle die klinisch bedeutsame repräsentiere, ist in der Literatur die Diskussion um die klinische Bedeutung der ASCUS-Befunde noch nicht abgeschlossen. Die Diagnose ASCUS beinhaltet eine weite Spanne von Veränderungen, über deren weiteres klinisches Verhalten nichts ausgesagt werden kann. Diese Spanne reicht von möglicherweise reaktiven über prämaligne bis hin zu möglicherweise malignen Veränderungen (Anton et al.; 2001). Auch liegen Hinweise darüber vor, dass, in Abhängigkeit von Alter/Menopausenstatus, möglicherweise vermehrt ASCUS-Befunde diagnostiziert werden, die sich bei Follow-up Kontrollen/Kontrollbiopsien als Plattenepitheldysplasien unterschiedlichen Ausmaßes darstellen lassen (Keating & Wang; 2001). Andererseits zeigt gerade die Gruppe der ASCUS-Befunde die höchste Variabilität bei der Reproduzierbarkeit der Befunde (Selvaggi; 2001). So wurden in einem großen Multicenter-RCT 38,6% der ASCUS-

Befunde bei erneuter zytologischer Musterung nachträglich als negativ eingestuft (Stoler & Schiffman; 2001).

Zusammenfassend muss also zum gegenwärtigen Zeitpunkt an der Forderung nach einer Testschwelle von HSIL+-Befunden festgehalten werden, eine Schwelle, bei der die bei Bergeron et al. (2001) ermittelten Testgütekriterien für AutoCyte PREP™ eine Verbesserung der Sensitivität bei gleichzeitiger Abnahme der Spezifität gegenüber Pap Tests zeigten, wenn für die zuerst verarbeiteten Proben getrennt ausgewertet wurde.

I C 6. 2. 3 Weitere methodische Aspekte

Ein besonderes Problem stellte die Zuordnung von Patienten zu konventionellen oder neuen Technologieverfahren dar. So ist z. B. bereits bei Broadstock (2000) darauf hingewiesen worden, dass Screeningtests, die mit einer finanziellen Mehrbelastung bei den Betroffenen verbunden sind, wie z.B. bei Dünnschichtpräparationen, zu einer Auswahlverzerrung führen können. Frauen mit vergleichsweise geringerem sozioökonomischem Status (bei gleichzeitig höherem Risiko für das Vorliegen eines Zervixkarzinoms) entscheiden sich seltener zu einer Testmethode mit zusätzlichen finanziellen Aufwendungen, als Frauen mit höherem sozioökonomischem Status (bei gleichzeitig geringerem Erkrankungsrisiko). Da in nur einer der Studien (Obwegeser et al.; 2001) ein Randomisationsverfahren eingesetzt wurde, kann bei den übrigen ein Selection-Bias dieser Art nicht ausgeschlossen werden.

Weiterhin machte Broadstock (2000) deutlich, dass die Einführung neuer Labormethoden mit einer erhöhten Aufmerksamkeit bei Untersuchern einhergeht, die zu einem erhöhten Auffinden pathologischer Befunde führt (diese Möglichkeit des Auftretens eines *Observer Bias* wird auch von Knottnerus et al. (2002) berichtet.) Besonders in diesem Zusammenhang kann aus einer erhöhten Detektionsrate nicht auf eine erhöhte Empfindlichkeit des Tests geschlossen werden. Bei Dünnschichtpräparationen besteht darüber hinaus noch eine andere Gefahr des Untersucher-Bias: wie bei Payne et al. (2000) und Broadstock (2000) beschrieben, unterliegen die Untersucher einer Lernkurve, die durch anfängliche Fehlinterpretation von Befunden mit Tendenz zum Upgrading gekennzeichnet ist und interindividuell unterschiedlich lange, z.T. über Monate anhalten kann, ein Problem, das z.B. bei Obwegeser et al. (2001) nicht auszuschließen ist.

Ebenfalls wurde bei Broadstock (2000) auf das Problem einer fehlenden Verblindungsmöglichkeit der Untersucher für die Art der Präparation (Dünnschicht vs. konventioneller Pap Test) hingewiesen, so dass Verzerrungen der Ergebnisse durch

subjektive Einstellungen der Untersucher zu diesen Methoden nicht ausgeschlossen werden können.

Als Einflussparameter für die berichteten Testgütekriterien stellte sich auch der Umgang mit inadäquaten Präparaten dar. So wurden bei Bergeron et al. (2001) die Befunde der inadäquaten Präparationen in die Berechnung von Sensitivität und Spezifität eingeschlossen, was zu größeren Differenzen zwischen Pap Test und Dünnschichtpräparaten hinsichtlich von Sensitivität und Spezifität führte. Nachträglich von uns an den berichteten Rohdaten durchgeführte Berechnungen unter Ausschluss genannter inadäquater Präparate zeigten teilweise sogar schlechtere Testgütekriterien für die Dünnschichtpräparate (Tabelle 24). Unserer Meinung nach stellt dieser Ansatz des Ausschlusses von inadäquaten Präparaten einen besseren Bezug zum Praxisalltag dar, da diese inadäquaten Präparate, die zytologisch nicht ausgewertet werden können, zu Wiederholungsuntersuchungen, nicht jedoch zu einer Diagnose führen. Die Bedeutung des Anteils an inadäquaten Präparaten (Adequacy) wird unter A.6.3.2 diskutiert.

Tabelle 24: Testgütekriterien mit/ohne Einschluss der inadäquaten Präparate

Parameter		Mit Einschluss der inadäquaten Präparate (wie berichtet in Bergeron et al. 2001)	Ohne Einschluss der inadäquaten Präparate (berechnet aus Rohdaten von Bergeron et al. 2001)
Sensitivität	Konv. Pap Test	82%	89%
	AutoCytePREP™	86%	86%
Spezifität	Konv. Pap Test	40%	52%
	AutoCyte™	43%	44%

I C 6. 3 Inhaltliche Aspekte

In diesem Kapitel werden die Ergebnisse der berücksichtigten Publikationen in bezug auf unsere eingangs gestellten Forschungsfragen diskutiert.

I C 6. 3. 1 Evidenzlage hinsichtlich diagnostischer Genauigkeit und Zuverlässigkeit

Dünnschichtpräparationstechniken

Wie bereits in der Diskussion der Methodik ausführlich dargelegt, konnten nur in einer Primärstudie Testgütekriterien wie Sensitivität, Spezifität, PPV, NPV und Likelihood-Ratios ermittelt werden (Bergeron et al.; 2001). Dort wurde eine leicht erhöhte Sensitivität für AutoCyte PREP™ Präparate im Vergleich zu konventionellen Pap Abstrichen beobachtet, gleichzeitig war die Spezifität jedoch etwas geringer. Diese Ergebnisse kehrten sich jedoch um, wenn in der Berechnung die inadäquaten Präparate nicht berücksichtigt wurden (siehe Tabelle 25). Auch ist im Abschnitt I A.6.2.1. ausführlich diskutiert worden, dass bei Untersuchung von Hochrisikokollektiven mit einer Überschätzung der Sensitivität des Tests zu rechnen ist (*Spectrum Bias*, hier auch zusätzlich *Selection Bias* nach Knottnerus et al.; 2002) Darüber hinaus beziehen sich die ermittelten Sensitivitäten nur auf einzelne Tests, eine Auswirkung auf das gesamte Screeningsystem mit jährlichen Untersuchungen kann daraus nicht sicher beurteilt werden.

Bezüglich der technischen Machbarkeit (Stufe 1 des Evaluationskonzepts von Flynn, 1996) haben Dünnschichtpräparationen in mehreren Studien zu einer Verbesserung der Adequacy geführt: Bergerano et al. (2001), Jensen et al. (2001). Dagegen beobachteten Obwegeser et al (2001) einen gegenläufigen Trend. Obwegeser et al. wiesen darauf hin, dass bei Einhaltung sorgfältiger Entnahmetechniken unter Verwendung adäquater Abstrichinstrumente eine bessere Abstrichqualität zu erzielen sei als mit Dünnschichtpräparationen und machten darauf aufmerksam, dass gerade die Adequacy durch Rückmeldungen des Zytologen an den Abstrichnehmer im Laufe ihrer Studie deutlich zu verbessern gewesen sei. Dies wird in der vorliegenden Arbeit jedoch nur am Rande berichtet, da diese Stufe der Evaluation bereits vorausgesetzt wurde und keine systematische Evaluation auf der Stufe der technischen Machbarkeit erfolgte. Es muss jedoch geschlussfolgert werden, dass auch auf dieser Stufe weiterer Forschungsbedarf zur Klärung bestehender Widersprüche besteht.

Computergestützte Auswertung von Pap Tests

Zu computergestützter Auswertung von Zervixabstrichen wurden nur zwei Studien identifiziert, beide Studien waren nicht randomisiert und erhielten bei inadäquaten Referenzstandards den Qualitätsscore D. Die bei Renshaw et al. (2000) berichteten FNR's dürfen bei fehlender histologischer Verifikation von mindestens 50% der hochgradig positiven Befunde nur mit Vorsicht interpretiert werden. Bereits in einer

Metaanalyse bei Arbyn & Schenck (2000) wurde auf die Unwirksamkeit von 10%-igem Rescreening negativer Befunde zur Senkung der FNR bei konventionellen Pap Tests hingewiesen, und es wurden wirksamere Verfahren, wie z.B. das schnelle Rescreening aufgezeigt. Die hier vorliegenden Publikationen haben keine Änderung der Evidenzlage im Vergleich zu den Berichten anderer HTA-Organisationen herbeigeführt, die diesen computerisierten Auswertungsverfahren keine Überlegenheit gegenüber konventionellen Pap Tests nachweisen konnten.

Zusammenfassung der Evidenzlage beider Technologien hinsichtlich diagnostischer Genauigkeit und Zuverlässigkeit

Bei Broadstock (2000) und McCrory et al. (1999) wurden insbesondere fehlende Angaben zur Spezifität neuer Technologien insgesamt, d.h. sowohl der Dünnschichtpräparationen als auch bei computergestützter Auswertung von Pap Tests betont. Nach Auswertung der berücksichtigten Publikationen ist die Evidenzlage zu Fragen diagnostischer Sicherheit durch neue Methoden unverändert schlecht. Fasst man Ergebnisse aus Primärstudien und Kontextdokumenten zusammen, sind alle hier bewerteten neuen Technologien eher etwas weniger spezifisch, d.h. es muss mit einer erhöhten Zahl falsch positiver Befunde gerechnet werden mit allen negativen Folgen für das Individuum und das gesamte Gesundheitssystem, die aus zusätzlichen, teilweise invasiven und mit Risiken behafteten Folgeuntersuchungen resultieren. Wie oben diskutiert, ist auch die Evidenzlage zur Sensitivität nicht ausreichend, um bei Anbietern oder Verbrauchern das Gefühl vermehrter Sicherheit zur Vermeidung falsch negativer Befunde zu implizieren.

Zum gegenwärtigen Zeitpunkt konnten von uns keine Studien identifiziert werden, die ohne wesentliche methodische Fehler unverzerrte Schätzungen für die Testgenauigkeit und Testzuverlässigkeit liefern konnten, weder im Bereich der Dünnschichtpräparationen noch bei der computerisierten Auswertung.

I C 6. 3. 2 Evidenzlage hinsichtlich prognostischer Indices

Es wurden keine prognostischen Primärstudien identifiziert. In den Kontextdokumenten konnten hierzu anhand der dort referierten Publikationen ebenfalls keine Aussagen gemacht werden. Somit fehlt die wissenschaftliche Evidenz über Vorteile der hier bewerteten neuen Technologien hinsichtlich Senkung von Inzidenz und Mortalität des Zervixkarzinoms gegenüber konventionellen Pap Tests.

I C 6. 3. 3 Übertragbarkeit internationaler Forschungsergebnisse

Externe Validität der Studienergebnisse

Zur Frage der Übertragbarkeit von Studienergebnissen, d.h. der externen Validität ihrer Ergebnisse ist bereits schon im Kapitel A.6.2. zur Diskussion methodischer Aspekte hingewiesen worden. Zusammenfassend sollen hier noch einmal wesentliche Kriterien genannt werden, an denen die externe Validität einzuschätzen ist (nach de Vet et al.; 2001):

- Studienpopulation: Eignung der Studienpopulation, Beschreibung von Ein- und Ausschlusskriterien, Beschreibung demographischer und klinischer Charakteristika der Studienpopulation, Beschreibung der Quelle der Studienpopulation;
- Diagnostischer Test und Referenzstandard: adäquate Definition eines Referenzstandards sowie der Testschwelle;
- Ergebnisse: Beschreibung der Ergebnisse in ausreichender Detailliertheit zur Berechnung der Testgütekriterien.

Keine der identifizierten Studien erfüllte alle diese Kriterien der externen Validität, insbesondere die Berichtsqualität der Studien war mit erheblichen Mängeln behaftet. Lijmer et al. (1999) fanden in ihrer bereits oben zitierten Arbeit empirische Evidenz dafür, dass die Genauigkeit von Tests z.B. auch in Studien mit mangelhafter Berichterstattung über die eingeschlossene Population überschätzt wurde.

Spezifika bei der Übertragbarkeit von Studienergebnissen im Rahmen des Zervixkarzinom-Screenings

Keine der identifizierten Publikationen stammte aus Deutschland. Da hinsichtlich der Krebsvorsorge gegen Gebärmutterhalskrebs prinzipiell in jedem Staat individuelle Programme existieren, sind die Ergebnisse nicht ohne Beachtung dieser Hintergründe übertragbar. Es waren unter den berücksichtigten Publikationen hinsichtlich ihres Herkunftslandes 1 französische, 1 US-amerikanische, 1 italienische, 1 dänische, 1 schweizerische und 1 südkoreanische Studie.

In den USA werden wie in Deutschland in jährlichem Abstand Abstrichuntersuchungen bei de facto opportunistischer Screeningstrategie durchgeführt. Die Prävalenzen sind jedoch in den USA sehr heterogen mit erheblichen Unterschieden zwischen verschiedenen ethnischen Bevölkerungsgruppen (SEER-Programm, NCI) und nicht ohne weiteres mit deutschen Erkrankungshäufigkeiten vergleichbar. Ebenso gibt es erhebliche Unterschiede zwischen beiden Ländern hinsichtlich allgemeinstaatlicher Kontrolle der Einhaltung von Qualitätsvorschriften und des Einflusses von Leitlinien auf Screening und klinisches Management. Aus diesem Grunde sind insbesondere Rescreening-Studien schlecht übertragbar. In den USA ist durch das gesetzlich vorgeschriebene 10% randomisierte Rescreening negativer Befunde und die Festsetzung einer maximal zulässigen Zahl von Präparaten pro Zytoassistent z.B. eine Alternative im Sinne eines Rapid Rescreening nicht durchsetzbar, da ein Zytoassistent bei dieser Methode rasch das gesetzliche Limit von Präparaten erreichen würde und für den restlichen Arbeitstag nicht mehr zur Verfügung stünde (Ruffing-Kullmann; 2001).

Frankreich führt, wie auch Dänemark, ein systematisches Screening in Dreijahresintervallen durch. Mit dieser Strategie werden in Dänemark über 90% der zu screenenden Frauen erreicht, also deutlich mehr als in Deutschland: hier werden auf ein dreijähriges Screeningintervall hochgerechnete Teilnahmeraten von etwa 80% angenommen (van Ballegooijen et al.; 2000). In Frankreich dagegen werden trotz systematischer Screeningstrategie ungenauen Schätzungen zufolge zwischen 22% und 69% der Frauen erreicht, Italien erreicht bei verschiedenartigen nebeneinander existierenden Strategien etwa 70% der Frauen bei dreijährigem Untersuchungsabstand (Linos & Ritza; 2000). Hinsichtlich der Erkrankungshäufigkeiten gehört beispielsweise Italien zu den 15 Staaten mit der weltweit geringsten Karzinominzidenz für Zervixkarzinome, gemeinsam mit einzelnen Schweizer Kantonen (Boyle et al.; 2000).

Systembezogene sowie epidemiologische Aspekte, wie die hier aufgeführten, beeinflussen zwar nicht die Testgütekriterien des einzelnen Screeningtests, wohl aber die Effektivität des Screeningprogramms unter Alltagsbedingungen.

I C 6. 4 Weiterer Forschungsbedarf

Aufgrund der hier ermittelten Evidenzlage ergibt sich nach Ansicht der Autoren der vorliegenden Arbeit weiterer Forschungsbedarf zur Prüfung der klinischen Effektivität neuer Screeningtechnologien:

1. Zunächst sollten klinische Studien zur Ermittlung unverzerrter Schätzungen für die diagnostische Genauigkeit und Zuverlässigkeit, wie z.B. der Testgütekriterien Sensitivität und Spezifität durchgeführt werden. Dabei ist ebenso auf die Auswahl einer geeigneten Studienpopulation, die Einhaltung eines akzeptablen Referenzstandards, wie auch auf designtypische Verzerrungsmöglichkeiten zu achten. Die Diskussion über einige Studiendesigns ist in der Literatur noch nicht abgeschlossen, wie z.B. beim Fall-Kontroll-Design, bei dem Lijmer et al. (1999) die größten aller von ihnen untersuchten Einflüsse auf die Verzerrung der Ergebnisse im Sinne einer Überschätzung der Leistungsfähigkeit eines Tests fanden, während andere Autoren das Fall-Kontroll-Design als mögliches Design diagnostischer Evaluationsstudien betrachten, um Testgüteparameter zu bestimmen (Knottnerus et al.; 2002). Die vorliegende Arbeit macht darüber hinaus auch deutlich, dass erheblicher weiterer Forschungsbedarf in der Entwicklung methodischer Standards zur Durchführung diagnostischer Studien besteht.
2. Nach Bestimmung der diagnostischen Genauigkeit und Zuverlässigkeit ist die Durchführung prospektiver Studien mit prognostischen Aussagen erforderlich, um die Effektivität des Gesamtscreeningprogramms zu beurteilen und den Impact neuer Technologien auf Inzidenz und Mortalität des Zervixkarzinoms einzuschätzen. Diese Studien sollten vorzugsweise in Deutschland durchgeführt werden, um die externe Validität bei im internationalen Maßstab nur schwer miteinander vergleichbaren Screeningprogrammen zu gewährleisten.
3. Obwohl zum Thema Screeningstrategie keine systematische Evaluation von Studien durchgeführt wurde, wurde deutlich, dass eine Durchführung von Studien zu Teilnahmeverhalten, Erreichbarkeit und Interventionsmöglichkeiten bei Hochrisikopatientinnen einen erfolversprechenden Ansatz für eine weitere Intensivierung des gesamten Screeningprogramms darstellt. Auch diese Studien sollten vorzugsweise in Deutschland durchgeführt werden, um die externe Validität zu gewährleisten.
4. Inwiefern eine Effektivitätssteigerung im Zervixkarzinom-Screening über eine Änderung von Qualitätskontrollen bei der Abstrichentnahme und im Laborprozess zu erreichen ist, ist im Rahmen von weiteren Untersuchungen zu bestimmen.
5. Im Ergebnis der Erkenntnis über die kausale Rolle bestimmter Hochrisikotypen des HPV-Virus bei der Zervixkarzinomentstehung sollten aus einer Technologiebewertung des Einflusses von HPV-Tests auf die Effektivität des Zervixkarzinom-Screenings neue Möglichkeiten aufgezeigt werden können.

I C 6. 5 Zusammenfassung der Diskussion - Effektivität

Aus den hier ausgewerteten Primärstudien und Kontextdokumenten konnte keine Evidenz für eine Überlegenheit Computergestützter Auswertungen oder Dünnschichtpräparationen gegenüber konventionellen Pap Tests gewonnen werden. Die diagnostische Genauigkeit und Zuverlässigkeit der neuen Technologien sind noch nicht hinreichend genau bestimmt worden, selbst auf der Stufe der technischen Machbarkeit (Flynn; 1996) bestehen noch widersprüchliche Forschungsergebnisse, aufgrund derer die Dünnschichtpräparationstechnik der konventionellen Technologie nicht als überlegen eingestuft werden kann.

Für die Computergestützte Auswertung von Pap Abstrichen mittels AutoPap™ liegen zwar Ergebnisse vor, die eine Überlegenheit als Rescreener gegenüber 10% manuellem Rescreening zur Senkung der Falsch-Negativen Rate zeigen, eine Überlegenheit gegenüber anderen Verfahren, wie z.B. dem manuellen *Rapid Rescreening* ist jedoch nicht belegt. Für den Einsatz von AutoPap™ als Primärscreener ergab sich ebenfalls keine neue Evidenz, aus der sich eine Überlegenheit der hier bewerteten Technologien hinsichtlich seiner diagnostischen Genauigkeit und Zuverlässigkeit hätte ableiten lassen.

Auch für Dünnschichtpräparationen konnte kein eindeutiger Vorteil hinsichtlich zu fordernder Testgütekriterien bewiesen werden. Die auf der Stufe technischer Machbarkeit anzusiedelnde *Adequacy* wird in der Literatur bei Dünnschichtpräparationstechniken wiederholt als gesteigert gegenüber konventionellen Pap Tests beschrieben. Die Hinweise für eine Steigerung der Adequacy waren jedoch nicht konsistent, wenn gleichzeitig die Sorgfalt bei der Entnahmetechnik konventioneller Pap Abstriche mit adäquaten Abstrichinstrumenten und Beachtung wichtiger Grundregeln einbezogen wurde.

I C 7 Schlussfolgerung

Aufgrund der bestehenden Evidenzlage konnte keine Überlegenheit der neuen Technologien, weder für Dünnschichtpräparationen noch für computergestützte Auswertung gegenüber konventionellen Pap Tests bezüglich ihrer diagnostischen Genauigkeit und Zuverlässigkeit demonstriert werden. Hier besteht weiterer Forschungsbedarf, auch dahingehend, inwiefern beide hier beschriebenen neuen Technologien in kombiniertem Einsatz eine Verbesserung der Testgütekriterien nach sich ziehen.

Die Anwendung der Dünnschichtpräparationen ergibt Präparate, die sich von den mit konventionellen Pap Tests gewonnenen Abstrichen unterscheiden. Die Dünnschichtpräparationen können gleichzeitig einen Informationsverlust und eine bessere Auswertbarkeit der Präparate bedeuten, wonach auch auf der Stufe technischer Machbarkeit (Flynn; 1996) noch keine klare Evidenzlage besteht.

Änderungen in der Screeningstrategie, im Screeningintervall oder in den Screeningteilnahmeraten können größere Auswirkungen auf die Effektivität des Screeningprogramms unter Alltagsbedingungen hervorrufen, als Verbesserungen der Sensitivität und Spezifität des einzelnen Tests.

I C 8 Literatur

I C 8. 1 Literatur allgemein

1. American Joint Committee on Cancer: AJCC Cancer Staging Manual, 5th Ed., Lippincott-Raven, Philadelphia-New York 1997.
2. Anonymous (1992). The Revised Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses: Report of the 1991 Bethesda Workshop. *The Journal of Reproductive Medicine* 37 (5), 383-386.
3. Anonymous (1998). Semi-automatic system for cancer screening cytology. *Gynakologe* 31 (2), 193-195.
4. Anton R.C., Ramzy I., Schwartz M.R., Younes P., Chakraborty S., Mody D.R. (2001). Should the Cytological Diagnosis of "Atypical Squamous Cells of Undetermined Significance" Be Qualified? An Assessment Including Comparison between Conventional and Liquid-Based Technologies. *Cancer Cytopathology* 93 (2), 93-99.
5. Anttila T., Saikku P., Koskela P., Bloigu A., Dillner J., Ikäheimo I., Jellum E., Lehtinen M., Lenner P., Hakulinen T., Närvänen A., Pukkala E., Thoresen S., Youngman L., Paavonen J. (2001). Serotypes of *Chlamydia trachomatis* and Risk for Development of Cervical Squamous Cell Carcinoma. *JAMA* 285 (1), 47-51.
6. Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland. (2002). Krebs in Deutschland. Häufigkeiten und Trends. 2., aktualisierte Ausgabe. Saarbrücken. <http://www.rki.de/KREBS> (Verantwortliche Autoren: Batzler W.U., Baumgardt-Elms C., Eisinger B., Lehnert M., Schön D., Schütz J., Stegmaier C.)
7. Arbyn M. & Schenck U. (2000). Detection of False Negative Pap Smears by Rapid Reviewing. A Metaanalysis. *Acta Cytol* 44 (6); 949-957.
8. ARD-Ratgeber Gesundheit, Sendetermin 18.10.1997; http://www.sfb.de/fernsehen/ratgeber/rgesundheit_arc.php3?idx=11
9. Ashfaq R., Liang Y., Saboorian M. H. (1995). Evaluation of PAPNET system for rescreening of negative cervical smears. *Diagn Cytopathol* 13 (1), 31-36.
10. (AWMF) A. d. W. M. F. (2000). Kurzgefasste Interdisziplinäre Leitlinien 2000: Zervixkarzinom. <http://www.uni-duesseldorf.de/WWW/AWMF/II/index.html>, 269ff.

11. Baltzer J., Meerpohl H. G., Bahnsen J., eds. (2000). *Praxis der Frauenheilkunde*, 2., aktualisierte Auflage edn. Thieme, Stuttgart, New York.
12. Banta D., Luce B. (1993). *Health technology and its assessment*. Oxford University Press, Oxford.
13. Barron B., Richart R. M. (1970). Statistical model of the natural history of cervical carcinoma. II: Estimates of the transition time from dysplasia to carcinoma in situ. *Journal of the National Cancer Institute* 45, 1025-1030.
14. Barron B. A., Richart R. M. (1968). A statistical model of the natural history of cervical carcinoma based on a prospective study of 557 cases. *Journal of the National Cancer Institute* 41, 1343-1353.
15. Barres D. & Bergeron C. (2000). [Reproducibility of cytologic diagnosis: study of CRISAP Ile-de-France]. *Gynecol Obstet Fertil* 28 (2); 120-126.
16. Beckmann M. W., Werner Y., Renner S. P., Fasching P. A., Jap D., Kuschel B. (2000). Krebsfrüherkennung in der Frauenärztlichen Praxis. Aktuelle Aspekte der wissenschaftlichen Diskussion. *Gynäkologe* 33 (7), 474-482.
17. Benedet J, Odicino F, Maisonneuve P, Severi G, Creasman W, Shepherd J, Sideri M and Pecorelli S. (1998). FIGO Staging of Gynecologic Cancer. *Carcinoma of the cervix uteri*, Page 1-3.
<http://www.figo.org/default.asp?id=/00000033.htm>;
<http://www.figo.org/default.asp?id=34>; <http://www.figo.org/default.asp?id=35>
18. Bethesda System 2001; www.bethesda2001.cancer.gov
19. Boyle P., Maisonneuve P., Autier P. (2000). Update on cancer control in women. *International Journal of Gynecology & Obstetrics* 70; 263-303.
20. Busse R, Orvain J, Velasco M, Perleth M et al.. (2002). Best practice in undertaking and reporting HTA. *Int J Technol Ass Health Care* 18 (2);361-422.
21. Canavan T. P., Doshi N. R. (2000). Cervical cancer. *Am Fam Physician* 61 (5), 1369-1376.
22. Cannistra S.A. & Niloff J.M. (1996). Cancer of the uterine cervix. *NEJM* 334 (16); 1030-1038.
23. Centre for Evidence-based Medicine; <http://www.minervation.com/cebm/>
24. Colgan T.J., Austin R.M.A., Davey D.D. (2001). The Annual Papanicolaou Test. Women's Safety and Public Policy. *Cancer Cytopathol* 93 (2); 81-85.
25. Cuzick J., Sasieni P., Davies P., Adams J., Normand C., Frater A., van Ballegooijen M., van den Akker E. (1999). A systematic review of the role of human papillomavirus testing within a cervical screening programme. In <http://www.ncchta.org/htapubs.htm>: The National Coordinating Centre for Health Technology Assessment (NCCHTA), pp. 204.

26. David H. (Hrs.) (1987). Wörterbuch der Medizin. VEB Verlag Volk und Gesundheit, Berlin. S. 1332.
27. Deeks JJ (2001). Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M; Smith GD; Altman DG (2001). Systematic reviews in health care. Metaanalysis in context. 2nd Edition. BMJ Books, BMJ Publishing Group, BMA House, London.
28. DeMay R.M. (1997). Common Problems in Papanicolaou Smear Interpretation. Arch Pathol Lab Med 121, 229-238.
29. De Vet HCW, van der Weijden T, Muris JWM, Heyrman J, Buntinx F, Knottnerus JA. (2001). European Journ of Epidemiol 17: 301-306.
30. Droste S. (2001). Ethische Aspekte des biochemischen Screenings. Vortrag auf 2.HTA-Symposium 3./4. Mai 2001, Boppard. Persönliche Mitteilung.
31. Dzuba IG; Díaz EY; Allen B; Leonard YF; Lazcano Ponce EC; Shah KV; Bishai D; Lorincz A; Ferris D; Turnbull B; Hernández Avila M; Salmerón J. The acceptability of self-collected samples for HPV testing vs. the pap test as alternatives in cervical cancer screening. Journal of women's health & gender-based medicine; 11; 3; 265-75; 2002
32. Engel J. & Hölzel D. Zur Epidemiologie des Zervixkarzinoms. In: Schenck, U., Seidl S., Schenck U.B. (Hrsg.). Referateband der 16. Fortbildungstagung Klinische Zytologie, München 29.11.-03.12.2001.
33. Engel J., Schmidt M., Schubert-Fritschle G., Tretter W., Hölzel D. (2000). Tumorregister München: Jahresbericht 1999 des klinisch-epidemiologischen Krebsregisters am Tumorzentrum München. Schwerpunkt: Gynäkologische Tumoren. W. Zuckschwerdt Verlag, München/Bern/Wien/New York.
34. Europäische Leitlinien für die Qualitätssicherung von Screening-Programmen zur Zervixkarzinomfrüherkennung. Vorgelegt im Rahmen des Programms der KEG GD V E.1. "Europa gegen Krebs". (1992). <http://www.zytologie.de/europais.htm>
35. Fahey M. T., Irwig L., Macaskill P. (1995). Meta-Analysis of Pap test accuracy. American Journal of Epidemiology 141 (7), 680-9.
36. FIGO. 1998. <http://www.figo.org>
37. Flynn & Adams 1996: Flynn KL, Adams EJ. Assessing diagnostic technologies. Boston: U.s. Department of Veterans Affairs Technology Assessment Programm (VATAP). Report No.1, 1996.
38. Food and Drug Administration. Internetseite des Archivs der bisher erteilten Pre Market Approvals: <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMA/pma.cfm> .
39. Gesundheitsbericht für Deutschland 1998, <http://www.gbe-bund.de/>

40. Gordis L (2001). Epidemiologie.. Dt. Erstausgabe. Übersetzt von Rüdiger Rau und Nicolaus Bocter.Kilian, Marburg.
41. Gupta D.K., Komaromy-Hiller G., Raab S.S., Nath M.E. (2001). Interobserver and Intraobserver Variability in the Cytologic Diagnosis of Normal and Abnormal Metaplastic Squamous Cells in Pap Smears. *Acta Cytol* 45 (5), 697-703.
42. Gustafsson L, Pontén J, Bergström R, Adami H-O. (1997). International Incidence Rates of Invasive Cervical Cancer Before Cytological Screening. *Int J Cancer* 71, 159-165.
43. Hillemanns P., Thaler C., Kimmig R. (1997). Epidemiology and diagnosis of cervical intraepithelial neoplasia--is the present concept of screening and diagnosis still current?. *Gynakol Geburtshilfliche Rundsch* 37 (4), 179-90.
44. Horten-Zentrum für praxisorientierte Forschung und Wissenstransfer, <http://www.evimed.ch/JournalClub/Glossar/glossar.asp?record=36>
45. Hutchinson M. L. (1996). Assessing the Costs and Benefits of Alternative Rescreening Strategies. *Acta Cytol* 40, 4-8.
46. Intersociety Working Group for Cytology Technologies. (1997). Proposed Guidelines for Primary Screening Instruments for Gynecologic Cytology. <http://www.acta-cytol.com/Admin/97may-guidelines.htm>
47. Intersociety Working Group for Cytology Technologies. (1998). A Proposed Methodology for Evaluating Secondary Screening (Rescreening) Instruments for Gynecologic Cytology. <http://www.cytology-iac.com/other/methodology.htm>
48. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. (2002) Designing studies to ensure that estimates of test accuracy are transferable. *BMJ* 324 (7338): 669-71.
49. Jepson R., Clegg A., Forbes C., Lewis R., Sowden A., Kleijnen J. (2000). The determinants of screening uptake and interventions for increasing uptake: a systematic review. *Health Technology Assessment NHS R & D HTA Programme*. Vol. 4; No. 14. <http://www.ncchta.org/htapubs.htm>
50. Jourda E.A. & Girardi F., Allgemeines Krankenhaus Wien (Universitätskliniken); <http://www.akhwien.ac.at/gyn/colposcopy/>
51. Keating J.T. & Wang H.H. Significance of a Diagnosis of Atypical Squamous Cells of Undetermined Significance for Papanicolaou Smears in Perimenopausal and Postmenopausal Women. *Cancer Cytopathology* 93 (2), 100-105.
52. Knottnerus JA, van Weel C, Muris JW. (2002). Evaluation of diagnostic procedures. *BMJ* 324 (7335): 477-80.
53. Koch J., Kirschner W., Schäfer A. Bestimmung der Prävalenz genitaler HPV- und Chlamydia-trachomatis-Infektionen in einem repräsentativen Querschnitt der weiblichen Normalbevölkerung in Berlin: *Infektionsepidmiol Forschung* 1997 (2), 1-7.

54. Köchli O., Sevin B., Benz J., Petru E., Haller U. (1998). Gynäkologische Onkologie: Manual für Klinik und Praxis. Springer, Berlin, Heidelberg, New York, Barcelona, Budapest, Hongkong, London, Mailand, Paris, Singapur, Tokio. pp. 149-161.
55. Kohlberger P.D., Stani J., Gitsch G., Kieback D.G., Breiteneker G. (1998). Comparative Evaluation of Seven Cell Collection Devices for Cervical Smears. *Acta Cytol* 43 (6), 1023-1026.
56. Kok M. R., Boon M. E. (2001). UITGAVE VAN VERENIGING HISTOTECHNIEK NEDERLAND ORGANISATIE CYTODIAGNOSTISCHE MEDEWERKERS. HISTOTECHNIEK CYTO-VISIE 9 (2); <http://www.pathweb.nl/>
57. Koss L. G. (1994). Reducing the error rate in Papanicolaou smears. One laboratory's experience with the PAPNET system. *Physician Assist* 18 (12), 48-52.
58. Küppers V., Schnurch H. G. (1997). Current state of early diagnosis of female genital cancer. *Gynakologe* 30 (8), 624-630.
59. Kunz R., Ollenschläger G., Raspe H.-G. (2000). Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis. Dt. Ärzte-Vlg., Köln, S. 413.
60. Kurman RJ, (ed.). (2002). Blaustein's pathology of the female genital tract. 5th Edit. New York : Springer. S. 325 – 346. ISBN: 0387952039.
61. Last J.M. (2001). A Dictionary of Epidemiology. 4th Edition. Oxford University Press.
62. Lee KR; Darragh TM; Joste NE; Krane JF; Sherman ME; Hurley LB; Allred EM; Manos MM. (2002). Atypical glandular cells of undetermined significance (AGUS): Interobserver reproducibility in cervical smears and corresponding thin-layer preparations. *Am J Clin Pathol*, 117 (1) 96-102.
63. Lee TH & Brennan TA. (2002). Direct-to-Consumer Marketing of High-Technology Screening Tests. *NEJM* 346 (7); 529-531.
64. Leidl R. (1998). Der Effizienz auf der Spur: Eine Einführung in die ökonomische Evaluation. In *Das Public Health Buch: Gesundheit und Gesundheitswesen*, Schwartz FW et al., ed. Urban & Schwarzenberg, pp. 346-369.
65. Leitlinie der AG Zervixpathologie und Kolposkopie. Leitlinien-Kurzfassung im Auftrag der Deutschen Krebsgesellschaft und der Deutschen Gesellschaft für Gynäkologie und Geburtshilfe. Diagnostische und therapeutische Standards bei intraepithelialen Neoplasien und frühinvasiven Karzinomen des unteren Genitaltraktes der Frau (ICD 10). Stand: 15.Juni 2000, München. <http://www.ag-cpc.de/leitlinien.html>
66. Leitlinie der Bundesärztekammer zur Qualitätssicherung zytologischer Untersuchungen im Rahmen der Früherkennung des Zervixkarzinoms (beschlossen vom Vorstand der BÄK am 17.12.1993; http://www.zytologie.de/leitlinie_der_bundesärztekammer_.htm

67. Levi F., Lucchini F., Negri E., Franceschi S., la Vecchia C. (2000). Cervical cancer mortality in young women in Europe: patterns and trends. *European Journal of Cancer* 36: 2266-2271.
68. Lijmer JG; Mol BW; Heisterkamp S; Bossel GJ; Prins MH; van der Meulen JH; Bossuyt PM. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 282(11): 1061-1066.
69. Linos A. & Riza E. (2000). Comparisons of cervical cancer screening programmes in the European Union. *European Journal of Cancer* 36; 2260-2265.
70. Martin-Hirsch P., Jarvis G., Kitchener H., Lilford R. (2000). *The Cochrane Library - 2000 Issue 2*. pp 1-17.
71. Martin-Hirsch P., Lilford R., Jarvis G., Kitchener H.C. (1999). Efficacy of cervical-smear collection devices: a systematic review and metaanalysis. *The Lancet* 354; 1763-1770.
72. Matuszewski K., Burnett D. (1996). *A guide to published technology assessments*. United Health Care Services Corporation, Oak Brook, Illinois, USA.
73. Medical College of Wisconsin; Developed by Hu C, Kneusel R, Barnas G. (1999). <http://www.intmed.mcw.edu/clincalc/bayes.html> .
74. Mitchell H & Medley G. (1995). Differences between Papanicolaou smears with correct and incorrect diagnoses. *Cytopathology*, 6(6):368-75.
75. NHSC (2000). *Opto-Electronic Devices for the Detection of Cervical Cancer*. New and Emerging Technologies Briefing Nr. 7.
76. Östör A. G. (1993). Natural history of cervical intraepithelial neoplasia: a critical review. *Int J GynecolPathol* 12 (2), 186-192.
77. Ouwerkerk-Noordam, E.; Boon, M.E.; Beck, S. (1994): Computer-assisted primary screening of cervical smears using the PAPNET method: comparison with conventional screening and evaluation of the role of the cytologist. *Cytopathology*, 5:4:211-218
78. Perleth M, Gerhardus A, Velasco-Garrido M (2003). *Positron-Emission-Tomography - Systematische Übersicht zur Wirksamkeit bei ausgewählten Indikationen*. Asgard-Verlag. Im Druck
79. Plaxe SC, Saltzstein SL. (1999). Estimation of the duration of the preclinical phase of cervical adenocarcinoma suggests that there is ample opportunity for screening. *Gynecol Oncol* 75(1): 55-61.
80. Raum E, Perleth M (2003). *Methoden zur Evaluation diagnostischer Tests*. Köln. Im Druck.
81. Richtlinien des Bundesausschusses der Ärzte und Krankenkassen über die Früherkennung von Krebserkrankungen („Krebsfrüherkennungsrichtlinien“, zuletzt geändert 25.05.1994). *Rechtsquellensammlung der KBV*, Stand 30.03.1998. <http://daris.kbv.de/>

82. Riethdorf L., Ramirez-Ponas J., Kühler-Obbarius C. (1999a). Diagnostik und Therapie zervikaler Plattenepitheldysplasien. *Pathologe* 20; 34-41.
83. Riethdorf L., Park T.W., Thomssen C. (1999b). Diagnostik und Therapie zervikaler Plattenepithelkarzinome. *Pathologe* 20; 42-49.
84. Robert Koch Institut. (1999). <http://www.rki.de/GBE/KREBS/KREBS.HTM>
85. Ruffing-Kullmann B., Remorini-Niedermeyer D., Schenck U., Arbyn M. (2001). Rapid Rescreening: Eine Alternative zur 10% Kontrolle. In: Schenck, U., Seidl S., Schenck U.B. (Hrsg.). Referateband der 16. Fortbildungstagung Klinische Zytologie, München 29.11.-03.12.2001.
86. Sackett DL, Haynes RB. (2002). The architecture of diagnostic research. *BMJ* 324 (7336): 539-41.
87. Schenck U. (2000) Zytologisches Vorsorgeprogramm - Durch neue Methoden ernsthaft herausgefordert? *Geburtsh Frauenheilk* 60 M125-M129.
88. Schenck U. (2001). Befundwiedergabe in der Zytologie: Münchner Nomenklatur II und Bethesda System 2001. In: Schenck, U., Seidl S., Schenck U.B. (Hrsg.). Referateband der 16. Fortbildungstagung Klinische Zytologie, München 29.11.-03.12.2001.
89. Schenck U. & Gassa K. (2001). Zertifizierung von Zytologieassistenten/innen durch die DGZ. In: Schenck, U., Seidl S., Schenck U.B. (Hrsg.). Referateband der 16. Fortbildungstagung Klinische Zytologie, München 29.11.-03.12.2001.
90. Schenck U., Moser B., Planding W., Pankau S. (2001). Implementation of Scanning Control in Visual Pap Smear Reading. (<http://www.zytologie.de/5thesacp.html>).
91. Schenck U., von Karsa L. (2000). Cervical cancer screening in Germany. *European Journal of Cancer* 36, 2221-2226.
92. Schneider A. (Hrsg.) (2001). Skriptum für die Hauptvorlesung "Frauenheilkunde". <http://www.uni-jena.de/ufk/cd/folder/start.htm>
93. Schneider A., Durst M., Jochmus I., Gissmann L. (1999). Epidemiology, aetiology and prevention of carcinoma of the cervix. *Gynakologe* 32 (4), 247-260.
94. Schneider A, Hoyer H, Dürst M. Bedeutung des Nachweises von humanen Papillomviren für die Vorsorge. *Dt Ärzteblatt* 2001;98: A 2517-2521.
95. Schneider A., Hoyer H., Lotz B., Leistritz S., Kuhne-Heid R., Nindl I., Müller B., Haerting J., Durst M. (2000). Screening for high-grade cervical intra-epithelial neoplasia and cancer by testing for high-risk HPV, routine cytology or colposcopy. *International Journal of Cancer* 89 (6), 529-34.
96. Schneider V. (1995). Qualitätssicherung in der Zytologie: Extrem-Beispiel U.S.A. *Der Frauenarzt Jahrg.* 36, 8/1995. 910-912.

97. Schneider V., Henry M.R., Jimenez-Ayala M., Turnbull L.S., Wright T.C. (2000). Cervical Cancer Screening, Screening Errors and Reporting. *Acta Cytol* 45 (4), 493-498.
98. SEER-Programm, NCI. (2002). Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Edwards BK (eds). SEER Cancer Statistics Review, 1973-1999, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1973_1999/, Tabelle I-15.
99. Selvaggi S.M. Implications of Low Diagnostic Reproducibility of Cervical Cytologic and Histologic Diagnosis. *JAMA* 285 (11), 1506-1508.
100. SGB V (2000) in Sozialgesetzbuch, Sonderausgabe für das Bundesministerium für Arbeit und Sozialordnung. Stand: 15. März 2000. Verlag C. H. Beck, München.
101. Sherlaw-Johnson C., Gallivan S., Jenkins D., Jones M. H. (1994). Cytological screening and management of abnormalities in prevention of cervical cancer: an overview with stochastic modelling. *JClinPathol* 47 (5), 430-435.
102. Slater D.N. (2000). Are women sufficiently well informed to provide valid consent for the cervical smear test? *Cytopathol* 11; 166-170.
103. Spitzer M. (1998). Cervical screening adjuncts: recent advances. *Am J Obstet-Gynecol* 179 (2), 544-556.
104. Stoler M.H. & Schiffman M. (2001). Interobserver Reproducibility of Cervical Cytologic and Histologic Interpretations. Realistic Estimates From the ASCUS-LSIL Triage Study. *JAMA* 285 (11), 1500-1505.
105. Svirbely J.R. & Sriram M.G. The Medical Algorithms Project. Chapter 39. Decisional Analysis. (<http://216.157.50.80/ftp/ch39.exe>; <http://216.157.50.80/index.html>; <http://216.157.50.80/ans.html#Q4>).
106. Tumorregister München. (1997). Daten zum Zervixkarzinom. http://www.krebsinfo.de/ki/daten/zervix/zervix_dat.html;
107. Tumorregister München. (1997). Epidemiologische Basisdaten: Zervixkarzinom. <http://www.krebsinfo.de/ki/daten/zervix/bas1zervix.html>;
108. Tumorregister München. (1997). Survival Zervixkarzinom. <http://www.krebsinfo.de/ki/daten/zervix/sur1zervix.html>
109. Tumorzentrum München (1999). Tumorzentrum München: Empfehlungen zur Diagnostik, Therapie und Nachsorge. Zervixkarzinom. Tumorzentrum München.
110. van Ballegooijen M., van den Akker-van Marle E., Patnick J., Lynge E., Arbyn M., Anttila A., Ronco G., Dik J., Habbema F. (2000). Overview of important cervical cancer screening process values in European Union (EU) countries, and tentative predictions of the corresponding effectiveness and cost-effectiveness. *European Journal of Cancer* 36; 2177-2188.

111. Vereinbarung zu Qualifikationsvoraussetzungen gemäß §135 Abs. 2 SGB V zur Durchführung von zytologischen Untersuchungen zur Diagnostik der Karzinome des weiblichen Genitale (letzte Fassung vom 01.10.1994). Rechtsquellensammlung der KBV, Stand 30.03.1998; <http://daris.kbv.de/>
112. WDR-Fernsehen, Sendung vom 26.03.2001. Gebärmutterhalskrebs (Zervixkarzinom): Früherkennung durch Virentest. Von Heiko Schmitz. http://www.wdr.de/tv/service/gesundheit/inhalte/010326_3.html.
113. Wulf KH & Schmidt-Matthiesen H (Hrsg.). (1991). Klinik der Frauenheilkunde und Geburtshilfe. Bd. 11. Spezielle gynäkologische Onkologie I. 3. Aufl. Urban & Schwarzenberg, München; Baltimore. S. 158 – 167.
114. Zentralinstitut für die Kassenärztliche Versorgung (1993). Krankheitsfrüherkennung Krebs, Männer und Frauen. Aufbereitung und Interpretation der Untersuchungsergebnisse aus den gesetzlichen Früherkennungsmaßnahmen 1989 und 1990. Deutscher Ärzte-Verlag, Köln.
115. Zollinger H.U. (1981). Pathologische Anatomie. Band II Spezielle Pathologie. 5. Aufl. Thieme, Stuttgart New York. S. 198 f.
116. Fa. Cytoc Germany GmbH i.Gr., Dr. Felix Dross, County Manager und Frau Anne Ritter, Marketing Manager, pers. Kommunikation 12.02.2002.
117. Fa. Medite GmbH, Generalvertrieb Deutschland für TriPath Imaging; pers. Kommunikation H. Goldbecker, 15.12.2001 und 15.01.2002.
118. Prof. Schenck, Zytologisches Institut der Bayerischen Krebsgesellschaft e.V.; Institut für Allgemeine Pathologie und Pathologische Anatomie der Technischen Universität München, persönliche Mitteilung 10.05.2001.
119. PD Dr. Volker Schneider, Angebot der Fa. Cytoc Corp. für ThinPREP™2000-Prozessor mit Listenpreisen und Vorzugspreisen und abnahmemengeabhängiger Preisstaffel der Testkits. Persönliche Kommunikation 06.07.2002.
120. Verband Cytologisch Tätiger Assistenten e.V. (VDCA), Frau Eva Müller-Leibinger, Vorsitzende; pers. Kommunikation 20.02.2002

I C 8. 2 Eingeschlossene Literatur

Eingeschlossene HTA-Berichte und Systematische Reviews

121. Australian Health Technology Advisory Committee (AHTAC): Review of automated and semi-automated cervical screening devices. AHTAC Publications approval No. 2416. Canberra: Commonwealth Dept. of Health and Family Services. April 1998. ISBN Nr. 0 642 36755 8
122. Bernstein SJ, Sanchez-Ramos L, and Ndubisi B.: Liquid-based cervical cytologic smear study and conventional Papanicolaou smears: A metaanalysis of prospective studies comparing cytologic diagnosis and sample adequacy. *Am J Obstet Gynecol* 2001; 185:308-317.
123. Broadstock M: Effectiveness and cost effectiveness of automated and semi-automated cervical screening devices: A systematic review of the literature. New Zealand Health Technology Assessment (NZHTA); The Clearing House for Health Outcomes and Health Technology Assessment NZHTA Report. 2000; 3(1).
124. McCrory, DC, Mather DB, Bastian L, et al.: Evaluation of Cervical Cytology. Evidence Report/Technology Assessment No. 5. (Prepared by the Duke University under Contract No. 290-97-0014) AHCPR Publication No. 99-E010. Rockville, MD: Agency for Health Care Policy and Research. February 1999.
125. Payne N, Chilcott J, McGoogan E: Liquid-based cytology in cervical screening: a rapid and systematic review. *Health Technol Assess* 2000; 4 (18).
126. Brown A.B. & Garber A.M. (1998). The cost-effectiveness of the three new technologies to enhance Pap testing. Chicago, ILL: Blue Cross & Blue Shield Association Technology Evaluation Center. In Broadstock, 2000.

Zitate aus HTA-Berichten, die bei der Technologiebewertung berücksichtigt wurden

127. Chock C., Irwig I., Berry G., Glasziou P. (1997). Comparing dichotomous screening tests when individuals negative on both tests are not verified. *Clin Epidemiol* 50: 1211-1217. In McCrory et al., 1999.
128. Corkill M., Knapp D., Martin J., Hutchinson M.L. (1997). Specimen adequacy of ThinPrep sample preparations in a direct-to-vial-study. *Acta Cytologica* 41: 39-44. In Broadstock, 2000.
129. Halford J. (1998). Cervical cytology – A review of evolving technologies. *Australian Journal of Medical Science* 19: 8-19. In Broadstock, 2000.

130. Nanda K., McCrory D.C., Myers E.R., Bastian L.A., Hasselblad V., Hickey J.D., Matchar D.B. (2000). Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Annals of Internal Medicine* 132: 810-819. In Broadstock et al., 2000.
131. Östor A. (1993). Natural history of cervical intraepithelial Neoplasia: a critical review. *International Journal of Gynecological Pathology* 12: 186-192. In Broadstock M; 2000.
132. Paul C., Bagshaw S., Bonita R., Durham G., Fitzgerald N.W., Jones R.W., Marshall B., and McAvoy B.R. (1991). Cervical screening recommendations: a working group report. *New Zealand Medical Journal* 104: 291-295. In Broadstock; 2000.
133. Reagan J.W., Fu Y.S. (1979). Histologic types and prognosis of cancers of the uterine cervix. *Int J Radiat Oncol Biol Phys* 5: 1015-1020. In Mc Crory et al., 1999.
134. Richart R.M. (1973). Cervical intraepithelial neoplasia. *Pathology Annual*. East Norwalk, CT: Appleton-Century-Crofts. P. 301-323. In Mc Crory et al., 1999.
135. Richart R.M. (1990). A modified terminology for cervical intraepithelial neoplasia. *Obstet Gynecol* 75(1):131-3. In Broadstock, 2000

Eingeschlossene Primärstudien

136. Alasio L.M., Alphandery C., Grassi P., Ruggeri M., De Palo G. and Pilotti S. (2001). Performance of the AutoPap Primary Screening System in the Detection of High-Risk Cases in Cervicovaginal Smears. *Acta Cytologica* 2001 ; 45; No.5: 704 – 708.
137. Bergeron C., Bishop J., Lemarie A., Cas F., Ayivi J., Huynh B., and Barrasso R. (2001). Accuracy of Thin-Layer Cytology in Patients Undergoing Cervical Cone Biopsy. *Acta Cytologica* 2001; 45; 519-524.
138. Jensen M.L., Fursted P.B., Svanholm H. (2001). Sammenligning af monolagsprover og konventionelle „smears“. *Ugeskr Laeger* 2001;163:1270-5.
139. Obwegeser J.H., and Brack S. (2001). Does Liquid Based Technology Really Improve Detection of Cervical Neoplasia? A Prospective, Randomized Trial Comparing the ThinPrep Pap Test with the Conventional Pap Test, Including Follow up of HSIL Cases *Acta Cytologica* 2001: 45; 709-714.
140. Park A., Lee S.N., Chae S.W., Park K.H., Kim J.W., and Lee H.P. (2001). Comparing the Accuracy of ThinPrep Pap Tests and Conventional Papanicolaou Smears on the Basis of the Histologic Diagnosis. A Clinical Study of Women with Cervical Abnormalities. *Acta Cytologica* 2001: 45; 525-531.

141. Renshaw A.A., Lezon K.M., and Wilbur D.C. (2001). The Human False-Negative rate of Rescreening Pap Tests. Measured in a Two-Arm Prospective Clinical Trial. *Cancer Cytopathology* 2001; 93: 106-110.
142. Fahs M.C., Plichta S.B., Mandelblatt J.S. (1996). Cost-effective policies for cervical cancer screening: An international review. *Pharmaco Economics* 9(3): 211-230. In CCOHTA-Report 1997.
143. Department of Health. First Report of the National Screening Committee. London, Dep. Of Health. (1998) In: Jepson et al.; 2000
144. General Medical Council. Seeking patients' consent: the ethical considerations. London. (1999). In: Jepson et al.; 2000
145. Olsen A.O., Gojen K., Sauer T. et al. (1995). Human papillomavirus and cervical intraepithelial neoplasia grade II-III: a population-based case-control study. *Int J Cancer* 61 (3): 312-315. In Cuzick et al., 1999
146. Schiffman M., Bauer H., Hoover R., et al. (1993). Epidemiologic evidence showing that HPV infection causes most cervical intraepithelial neoplasia. *J Natl Cancer Inst* 85 : 958-964. In Cuzick et al.; 1999

I C 8. 3 Ausgeschlossene Literatur

Ausgeschlossene HTA-Berichte (*Ausschlussgrund kursiv*)

147. Canadian Coordinating Office for Health Technology Assessment (CCOHTA): Assessment of Techniques for Cervical Cancer Screening. Report 1997: 2E. National Library of Canada. May 1997. ISBN 1-895561-49-3. *Damaliger Stand des Wissens ohne systematische Technologiebewertung referiert.*
148. Larizgoitia I. Diagnóstico precoz del cáncer cervical: conocimiento actual sobre viejas y nuevas tecnologías. Barcelona: Agència d'Avaluació de Tecnologia i Recerca Mèdiques. Servei Català de la Salut. Departament de Sanitat i Seguretat Social. Generalitat de Catalunya. Octubre 2001 (BR02/2001). *Kurz-HTA ohne zusätzlichen Informationsgewinn.*

Ausgeschlossene Primärstudien (*Ausschlussgrund kursiv*)

149. Boman F., Farre I., Marine M.O. et al. (1999). Pourquoi nous préférons la technique en couche mince aux frottis cervico-utérins conventionnelles. *Clin Exp Path* 1999;47:81-87. *Studie zur technischen Machbarkeit, keine Outcomeparameter zur diagnostischen Genauigkeit bestimmt.*

150. Marino J.F., and Fremont-Smith M. (2001). Direct-to-Vial Experience with AutoCyt[®] PREP in a Small New England Regional Cytology Practice. *The Journal of Reproductive Medicine* 2001; 46: 353-358. *Studie zur technischen Machbarkeit, keine Outcomeparameter zur diagnostischen Genauigkeit bestimmt.*
151. Massarani-Wafai et al.: Evaluation of Cellular Residue in the ThinPrep[™] PreservCyt[™] Vial. *Diagnostic Cytopathology* 2000; 23; No. 3: 208-212. *Studie zum Benefit eines zusätzlichen Abstriches, kein Outcomeparameter gemäß Einschlusskriterien.*
152. Monsonego J., Autillo-Touati A., Bergeron C., Dachez R., Liaras J., Saurel J., Zerat L., Chatelain P., and Mottot C. (2001). Liquid-based cytology for primary cervical cancer screening: a multi-centre study. *British Journal of Cancer* 2001; 84(3): 360-366. *Studie zur technischen Machbarkeit, keine Outcomeparameter zur diagnostischen Genauigkeit bestimmt.*
153. Shield et al.: Improving Cervical cytology screening in a remote, high risk population. *MJA* 1999; 170: 255-258. *Studie zur technischen Machbarkeit, keine Outcomeparameter zur diagnostischen Genauigkeit bestimmt.*
154. Tench W. Preliminary assessment of the AutoCyt[®] PREP. Direct-to-vial performance. *J Reprod Med* 2000;45:912-6. *Studie zur technischen Machbarkeit, keine Outcomeparameter zur diagnostischen Genauigkeit bestimmt.*

I Anhang

I 1 Vorsorgeschein

AOK	LKK	BKK	IKK	VdAK	AEV	Knappschaft
Name, Vorname des Versicherten						
geb. am						
Kassen-Nr.		Versicherten-Nr.		Status		
Vertragsarzt-Nr.		VK gültig bis		Datum		

Geburtsjahr: Tag der Untersuchung:

Anamnese

Wurde bereits eine Krebsfrüherkennungsuntersuchung durchgeführt? nein 31 ja 32

Wenn ja: Zuletzt im Jahr (Endziffern eintragen) 32 33

Ggf. Nr. des letzten zyt. Bef.: _____ Gruppe: _____

Gyn. Op., Strahlen- oder Chemotherapie des Genitales: nein 34 ja 35

welche? wann? _____

Zahl der Schwangerschaften einschl. Fehlgeburten 35 36

Jetzt: _____

Letzte Periode: _____ nein 1 ja 2

Gravidität: _____ 37

Path. gyn. Blutungen: _____ 38

z.B.: Zwischen den normalen Regeln, Dauer- oder Schmierblutung im Klimakterium, in der Postmenopause, bei Verkehr, blutig-bräunlicher Ausfluß

Sonstiger Ausfluß: _____ 1 2 39

IUP: _____ 40

Ovulationshemmer: _____ 41

Sonstige Hormon-Anwendung: _____ 42

welche? warum? _____

Befund

Vulva: _____ nein ja 43

Inspektion auffällig: _____ 43

Portio und Vagina: _____

Spiegeleinstellung auffällig: _____ 44

Inneres Genitale: _____

Gyn. Tastbefund auffällig: _____ 45

Inguinale Lymphknoten auffällig: _____ 46

Bish. unbek. behandlungsbed. Nebenbefunde: _____ 47

Haut:	Wachstum, Verfärbung oder Blutung eines Pigmentfleckens oder Knotens	nein <input type="checkbox"/> 7 ja <input type="checkbox"/> 8	Zusätzl. ab 30. Lebensjahr
Mamma:	Inspektions-/Tastbefund auffällig	<input type="checkbox"/> 49	Zusätzl. ab 45. Lebensjahr
	Axilläre Lymphknoten auffällig	<input type="checkbox"/> 50	
Rektum/Kolon:	Abgang von Blut oder Schleim mit dem Stuhl	nein <input type="checkbox"/> 1 ja <input type="checkbox"/> 2	Zusätzl. ab 45. Lebensjahr
	Neu aufgetr. Unregelmäßigkeiten im Stuhlgang	<input type="checkbox"/> 51	
	Tastbefund auffällig	<input type="checkbox"/> 52	
	Stuhlttest zurückgegeben	<input type="checkbox"/> 53	
	Stuhlttest positiv	<input type="checkbox"/> 54	

Krebsfrüherkennung – Frauen

Zytologischer Befund	
Unters.-Nr. _____	Eingangsdatum: _____
Endo-zervikale Zellen	Ausgangsdatum: _____
Gruppe I/II	Gruppe IV a
Gruppe III	Gruppe IV b
Gruppe III D	Gruppe V
Döderleinflora	Mischflora
Trichomonaden	Mykose
	Kokkenflora
	Aminkolpitis
Proliferationsgrad: _____	
Empfehlung:	
zytologische Kontrolle	nach Entzündungsbehandlg.
histologische Klärung	nach Östrogenbehandlung
Bemerkungen:	
Arztstempel	Unterschrift des zytologisch tätigen Arztes

Gyn. Diagnose _____

Weitere Diagnostik wegen Krebsverdachts erforderlich: nein <input type="checkbox"/> 59 ja <input type="checkbox"/> 60			
falls Krebsverdacht: _____			
	Ergebnis		
Krebsverdacht bei:	Histologisch gesicherte(s)	Verd. nicht bestät.	Diagnostik nicht abgeschlossen*
Mamma	Mamma-Ca		
Cervix uteri	Dysplasie, l. - m.		
Corpus uteri	Dysplasie, s./CIS		
Übr. Genitale	invasives Ca		
Rektum/Kolon	Corpus-Ca		
Haut	Ca des übr. Genit.		
	Rektum-Ca		
	Kolon-Ca		
	große(s) Adenom(e) (≥ 10 mm)		
	kleine(s) Adenom(e) (< 10mm)		
	Metanom		
	sonst. Malignom der Haut		
RR	Grund: Patientin entzog sich weiterer Diagnostik		
(bei Werten über 140/90 bitte 2. Messung eintragen)	Lymphknotenbefall	Abrechnungszeitpunkt erreicht *	
	nein <input type="checkbox"/> 72 ja <input type="checkbox"/> 73	<input type="checkbox"/> 74	
	Diagnose _____		

* bei ausstehendem Ergebnis bitte mindestens ein Quartal abwarten!

Ausfertigung I. untersuch. Arzt

Vertragsarztstempel

Unterschrift

I 2 Dokumentation der Recherchen

Übersicht 1. Berücksichtigte Einrichtungen

ACPM	American College of Preventive Medicine
AHCPR	Agency for Health Care Policy and Research (USA)
AHTAC	Australian Health Technology Advisory Committee
AMA	Alberta Medical Association
ANDEM	Agence Nationale pour le Développement de l'Évaluation Médicale (Frankreich)
ASCCP	American Association of Colposcopy and Cervical Pathology (USA)
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
BÄK	Bundesärztekammer
CDHSH	Commonwealth Department of Human Services and Health (Australien)
CMA	Canadian Medical Association
CTFPHC	Canadian Task Force on Preventive Health Care
DGZ	Deutsche Gesellschaft für Zytologie
FDA	Food and Drug Administration (USA)
ICSI	Institute for Clinical Systems Improvement (USA)
INAHTA	International network of Agencies for HTA
ISWGCT	Intersociety Working Group for Cytology Technologies (USA)
NCI	National Cancer Institute (USA)
NCCHTA	National Coordinating Centre for Health Technology Assessment (UK)
NHMRC	National Health and Medical Research Council (Australien)
NICE	National Institute for Clinical Excellence (UK)
NIH	National Institute of Health (USA)
NZHTA	New Zealand Health Technology Assessment
OTA	Office of Technology Assessment (USA)
SGOC	Society of Gynecologic Oncologists of Canada
USPSTF	US Preventive Task Force (USA)

Übersicht 2: Suchstrategie der systematischen Literaturrecherche klinische Effektivität

Suche	Ergebnisse
#53 #45 or #52	393
#52 #24 and #51	184
#51 #46 or #47 or #48 or #49 or #50	20701
#50 'Randomized-Controlled-Trials' / all subheadings in MIME, MJME	3506
#49 explode 'Sensitivity-and-Specificity' / all subheadings in MIME, MJME	16301
#48 'Evaluation-Studies' / all subheadings in MIME, MJME	628
#47 explode 'Review-Literature' / all subheadings in MIME, MJME	181
#46 'Meta-Analysis' / all subheadings in MIME, MJME	532
#45 #25 or #26 or #27 or #28 or #29 or #30 or #31 #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44	269
#44 #24 and (consensus-development-conference-NTH in PT)	0
#43 #24 and (consensus-development-conference in PT)	4
#42 #24 and (review-academic in PT)	5
#41 #24 and (review-literature in PT)	3
#40 #24 and (practice-guideline in PT)	8
#39 #24 and (randomized-controlled-trial in PT)	22
#38 #24 and (randomized near trial)	9
#37 #24 and (clinical near trial)	10
#36 #24 and blind	10
#35 #24 and efficien*	20
#34 #24 and efficac*	40
#33 #24 and effectiv*	116
#32 #24 and (meta-analysis in PT)	6
#31 #24 and metaanaly*	2
#30 #24 and meta-analy*	4
#29 #24 and (evaluat* near study)	41
#28 #24 and report	42
#27 #24 and overview	3
#26 #24 and (guideline in PT)	11
#25 #24 and guideline	8
#24 #23 not (case-report in TG)	923
#23 #22 and (human in TG)	986
#22 #9 or #15 or #19 or #21	1057
#21 #17 and #20	165
#20 #10 or #13	1418
#19 #17 and #18	511
#18 #3 or #4 or #5 or #6 or #7 or #8 or #11 or #12 or #14 or #16	37152
#17 #1 or #2	1336
#16 rescreening	19
#15 (semiautomat* or automat* or method*) and (screen* or smear) and (cervi* or vagina*)	394
#14 autocyte	3
#13 (semiautomat* or automat* or rapid) and (screen* or smear*)	1157
#12 cytorich	1
#11 thinprep	34
#10 (pap or papanicolaou or papnet) and (screen* or smear*)	273
#9 (vagina* or cervi*) and (screen* or smear*)	942
#8 'Cytodiagnosis-' / all subheadings in MIME, MJME	155
#7 'Cytology-' / all subheadings in MIME, MJME	34
#6 explode 'Cytological-Techniques' / all subheadings in MIME, MJME	29737
#5 'Image-Processing-Computer-Assisted' / all subheadings in MIME, MJME	3236
#4 explode 'Mass-Screening' / all subheadings in MIME, MJME	4217
#3 'Vaginal-Smears' / all subheadings in MIME, MJME	504
#2 'Cervical-Intraepithelial-Neoplasia' / all subheadings in MIME, MJME	279
#1 'Cervix-Neoplasms' / all subheadings in MIME, MJME	1303
Ergebnisse von: #45 or #52	

I 3 Checklisten: HTA-Berichte und Systematische Übersichtsarbeiten

Titel: Review of automated and semi-automated cervical screening devices.					
Autoren: Australian Health Technology Advisory Committee					
Quelle: Canberra: Commonwealth Dept. of Health and Family Services. April 1998. ISBN 0 642 367558					
Dokumenttyp: Assessment Dokument					
Adressaten: Entscheidungsträger X Kliniker X Patienten X Andere					
Klas	A. Fragestellung und Kontext	Ja	Nein	?	
I	Werden Anlass und Ziel der Publikation im Sinne einer „Policy Question“ dargestellt?	x			
QA	Gibt es im Rahmen des breiteren Kontext eine präzise formulierte Forschungsfrage nach der (interessierenden) Intervention?	x			
I	Sind in der Publikation Angaben zu folgenden Aspekten enthalten:				
I	Epidemiologie der Zielerkrankung	x			
I	(Entwicklungs-)stand der Technologie	x			
I	Efficacy	x			
I	Effectiveness	x			
I	Nebenwirkungen				x
I	Indikationen				x
I	Kontraindikationen				x
I	Praxisvariation				x
I	Versorgungsstrukturen	x			
I	Kostengesichtspunkten	x			
I	sozioökonomischen, ethischem und juristischen Impact	x			
Klas	B. Methodik der Informationsgewinnung				
QA	Wurden die genutzten Quellen dokumentiert?	x			
QB	Wurden die Suchstrategien dokumentiert?	x			
QB	Wurden Einschlusskriterien definiert?	x			
QA	Wurden Ausschlusskriterien definiert?		x		
Klas	C. Methodik der Bewertung der Dokumentation				
QA	Wurden Validitätskriterien berücksichtigt?	x			
QC	Wurde die Bewertung unabhängig von mehreren Personen durchgeführt?		x		
QC	Sind ausgeschlossene Studien mit ihrem Ausschlussgründen dokumentiert?		x		
QC	Ist die Datenextraktion nachvollziehbar dokumentiert?	x			
QC	Erfolgte die Datenextraktion von mehreren Personen unabhängig?		x		
Klas	D. Methodik der Informationssynthese				
I	Es wurden quantitative Informationssynthesen durchgeführt: (Metaanalyse-Bogen 1b)		x		
I	Es wurden qualitative Informationssynthesen durchgeführt: (Bogen 1b)	x			
I	Es wurden zur Ergänzung der Datenlage eigene Erhebungen durchgeführt:		x		
Klas	E. Ergebnisse/Schlussfolgerungen				
QB	Wird die bestehende Evidenz in den Schlussfolgerungen konsequent umgesetzt?	x			
QA	Werden methodisch bedingte Limitationen der Aussagekraft kritisch diskutiert?	x			
I	Werden Handlungsempfehlungen ausgesprochen?	x			
I	Gibt es ein Grading der Empfehlungen?				x
QC	Wurde die Publikation vor der Veröffentlichung einem externen Reviewverfahren unterzogen?				x
I	Ist ein „Update“ der Publikation eingeplant?				x
Klas	F. Übertragbarkeit der internationalen/ausländischen Ergebnisse und Schlussfolgerungen				
	Bestehen Unterschiede hinsichtlich der / des:				
	Epidemiologie der Zielkondition?		x		
	Entwicklungsstandes der Technologie?		x		
	Indikationsstellung?		x		
	Versorgungskontexte, -bedingungen, -prozess?	x			
	Vergütungssysteme?				x
	Sozioökonomischen Konsequenzen?	x			
	Patienten- und Providerpräferenzen?		x		
Beurteilung: Berücksichtigt x					

Titel: Effectiveness and cost effectiveness of automated and semi-automated cervical screening devices.				
Autoren: Broadstock, Marita				
Quelle: NZHTA Report 2000; 3(1)				
Dokumenttyp: Assessment Dokument				
Adressaten: Entscheidungsträger Kliniker X Patienten Andere				
Klas	A. Fragestellung und Kontext	Ja	Nein	?
I	Werden Anlass und Ziel der Publikation im Sinne einer „Policy Question“ dargestellt?	x		
QA	Gibt es im Rahmen des breiteren Kontext eine präzise formulierte Forschungsfrage nach der (interessierenden) Intervention?	x		
I	Sind in der Publikation Angaben zu folgenden Aspekten enthalten:			
I	Epidemiologie der Zielerkrankung			x
I	(Entwicklungs-)stand der Technologie	x		
I	Efficacy	x		
I	Effectiveness	x		
I	Nebenwirkungen	x		
I	Indikationen	x		
I	Kontraindikationen			x
I	Praxisvariation			x
I	Versorgungsstrukturen	x		
I	Kostengesichtspunkten	x		
I	sozioökonomischen, ethischem und juristischen Impact	x		
Klas	B. Methodik der Informationsgewinnung			
QA	Wurden die genutzten Quellen dokumentiert?	x		
QB	Wurden die Suchstrategien dokumentiert?	x		
QB	Wurden Einschlusskriterien definiert?	x		
QA	Wurden Ausschlusskriterien definiert?	X		
Klas	C. Methodik der Bewertung der Dokumentation			
QA	Wurden Validitätskriterien berücksichtigt?	x		
QC	Wurde die Bewertung unabhängig von mehreren Personen durchgeführt?		x	
QC	Sind ausgeschlossene Studien mit ihrem Ausschlussgründen dokumentiert?		x	
QC	Ist die Datenextraktion nachvollziehbar dokumentiert?			x
QC	Erfolgte die Datenextraktion von mehreren Personen unabhängig?		x	
Klas	D. Methodik der Informationssynthese			
I	Es wurden quantitative Informationssynthesen durchgeführt: (Metaanalyse-Bogen 1b)		x	
I	Es wurden qualitative Informationssynthesen durchgeführt: (Bogen 1b)	x		
I	Es wurden zur Ergänzung der Datenlage eigene Erhebungen durchgeführt:		x	
Klas	E. Ergebnisse/Schlussfolgerungen			
QB	Wird die bestehende Evidenz in den Schlussfolgerungen konsequent umgesetzt?	x		
QA	Werden methodisch bedingte Limitationen der Aussagekraft kritisch diskutiert?	x		
I	Werden Handlungsempfehlungen ausgesprochen?	x		
I	Gibt es ein Grading der Empfehlungen?			x
QC	Wurde die Publikation vor der Veröffentlichung einem externen Reviewverfahren unterzogen?	x		
I	Ist ein „Update“ der Publikation eingeplant?	x		
Klas	F. Übertragbarkeit der internationalen/ausländischen Ergebnisse und Schlussfolgerungen			
	Bestehen Unterschiede hinsichtlich der / des:			
	Epidemiologie der Zielkondition?			x
	Entwicklungsstandes der Technologie?			x
	Indikationsstellung?			x
	Versorgungskontexte, -bedingungen, -prozess?			x
	Vergütungssysteme?			x
	Sozioökonomischen Konsequenzen?			x
	Patienten- und Providerpräferenzen?			x
Beurteilung: Berücksichtigt: Ja				

Titel: Evaluation of Cervical Cytology; Evidence Report/Technology Assessment No. 5 (Prepared by Duke Univ. under Contract No. 290-97-0014)				
Autoren: McCrory, DC, Mather, DB, Bastian, L, et al.				
Quelle: AHCPR Publication No. 99-E010; Rockville, MD: Agency for Health Care Policy and Research; February 1999				
Dokumenttyp: Evidence Report/Technology Assessment No. 5				
Adressaten: Entscheidungsträger Kliniker X Patienten Andere				
Klas	A. Fragestellung und Kontext	Ja	Nein	?
I	Werden Anlass und Ziel der Publikation im Sinne einer „Policy Question“ dargestellt?	x		
QA	Gibt es im Rahmen des breiteren Kontext eine präzise formulierte Forschungsfrage nach der (interessierenden) Intervention?	x		
I	Sind in der Publikation Angaben zu folgenden Aspekten enthalten:			
I	Epidemiologie der Zielerkrankung	x		
I	(Entwicklungs-)stand der Technologie	x		
I	Efficacy	x		
I	Effectiveness	x		
I	Nebenwirkungen	x		
I	Indikationen	x		
I	Kontraindikationen			x
I	Praxisvariation			
I	Versorgungsstrukturen		x	
I	Kostengesichtspunkten	x		
I	sozioökonomischen, ethischem und juristischen Impact	x		
Klas	B. Methodik der Informationsgewinnung			
QA	Wurden die genutzten Quellen dokumentiert?	x		
QB	Wurden die Suchstrategien dokumentiert?	x		
QB	Wurden Einschlusskriterien definiert?	x		
QA	Wurden Ausschlusskriterien definiert?	x		
Klas	C. Methodik der Bewertung der Dokumentation			
QA	Wurden Validitätskriterien berücksichtigt?	x		
QC	Wurde die Bewertung unabhängig von mehreren Personen durchgeführt?	x		
QC	Sind ausgeschlossene Studien mit ihrem Ausschlussgründen dokumentiert?		x	
QC	Ist die Datenextraktion nachvollziehbar dokumentiert?	x		
QC	Erfolgte die Datenextraktion von mehreren Personen unabhängig?	x		
Klas	D. Methodik der Informationssynthese			
I	Es wurden quantitative Informationssynthesen durchgeführt: (Metaanalyse-Bogen 1b)	x		
I	Es wurden qualitative Informationssynthesen durchgeführt: (Bogen 1b)	x		
I	Es wurden zur Ergänzung der Datenlage eigene Erhebungen durchgeführt:		x	
Klas	E. Ergebnisse/Schlussfolgerungen			
QB	Wird die bestehende Evidenz in den Schlussfolgerungen konsequent umgesetzt?	x		
QA	Werden methodisch bedingte Limitationen der Aussagekraft kritisch diskutiert?	x		
I	Werden Handlungsempfehlungen ausgesprochen?	x		
I	Gibt es ein Grading der Empfehlungen?		x	
QC	Wurde die Publikation vor der Veröffentlichung einem externen Reviewverfahren unterzogen?	x		
I	Ist ein „Update“ der Publikation eingeplant?		x	
Klas	F. Übertragbar. der internationalen/ausländischen Ergebnisse u. Schlussfolgerungen	x		
	Bestehen Unterschiede hinsichtlich der / des:	x		
	Epidemiologie der Zielkondition?	x		
	Entwicklungsstandes der Technologie?	x		
	Indikationsstellung?	x		
	Versorgungskontexte, -bedingungen, -prozess?		x	
	Vergütungssysteme?		x	
	Sozioökonomischen Konsequenzen?	x		
	Patienten- und Providerpräferenzen?			x
Beurteilung: Berücksichtigt: Ja				

Titel: Liquid-based cytology in cervical screening: a rapid and systematic review				
Autoren: Payne N., Chilcott J., McGoogan E.				
Quelle: Health Technology Assessment NHS R&D HTA Programme. 2000; 4(18)				
Dokumenttyp: Assessment Dokument				
Adressaten: Entscheidungsträger X Kliniker X Patienten Andere				
Klas	A. Fragestellung und Kontext	Ja	Nein	?
I	Werden Anlass und Ziel der Publikation im Sinne einer „Policy Question“ dargestellt?	x		
QA	Gibt es im Rahmen des breiteren Kontext eine präzise formulierte Forschungsfrage nach der (interessierenden) Intervention?	x		
I	Sind in der Publikation Angaben zu folgenden Aspekten enthalten:			
I	Epidemiologie der Zielerkrankung	x		
I	(Entwicklungs-)stand der Technologie	x		
I	Efficacy	x		
I	Effectiveness	x		
I	Nebenwirkungen			
I	Indikationen	x		
I	Kontraindikationen			
I	Praxisvariation	x		
I	Versorgungsstrukturen	x		
I	Kostengesichtspunkten	x		
I	sozioökonomischem, ethischem und juristischem Impact			
Klas	B. Methodik der Informationsgewinnung			
QA	Wurden die genutzten Quellen dokumentiert?	x		
QB	Wurden die Suchstrategien dokumentiert?	x		
QB	Wurden Einschlusskriterien definiert?	x		
QA	Wurden Ausschlusskriterien definiert?	x		
Klas	C. Methodik der Bewertung der Dokumentation			
QA	Wurden Validitätskriterien berücksichtigt?	x		
QC	Wurde die Bewertung unabhängig von mehreren Personen durchgeführt?		x	
QC	Sind ausgeschlossene Studien mit ihren Ausschlussgründen dokumentiert?		x	
QC	Ist die Datenextraktion nachvollziehbar dokumentiert?	x		
QC	Erfolgte die Datenextraktion von mehreren Personen unabhängig?		x	
Klas	D. Methodik der Informationssynthese			
I	Es wurden quantitative Informationssynthesen durchgeführt: (Metaanalyse-Bogen 1b)		x	
I	Es wurden qualitative Informationssynthesen durchgeführt: (Bogen 1b)	x		
I	Es wurden zur Ergänzung der Datenlage eigene Erhebungen durchgeführt:		x	
Klas	E. Ergebnisse/Schlussfolgerungen			
QB	Wird die bestehende Evidenz in den Schlussfolgerungen konsequent umgesetzt?	x		
QA	Werden methodisch bedingte Limitationen der Aussagekraft kritisch diskutiert?	x		
I	Werden Handlungsempfehlungen ausgesprochen?	x		
I	Gibt es ein Grading der Empfehlungen?		x	
QC	Wurde die Publikation vor der Veröffentlichung einem externen Reviewverfahren unterzogen?	x		
I	Ist ein „Update“ der Publikation eingeplant?	x		
Klas	F. Übertragbarkeit der internationalen/ausländischen Ergebnisse und Schlussfolgerungen			
	Bestehen Unterschiede hinsichtlich der / des:			
	Epidemiologie der Zielkondition?	x		
	Entwicklungsstandes der Technologie?		x	
	Indikationsstellung?	x		
	Versorgungskontexte, -bedingungen, -prozess?	x		
	Vergütungssysteme?	x		
	Sozioökonomischen Konsequenzen?		x	
	Patienten- und Providerpräferenzen?	x		
Beurteilung: Berücksichtigt				

Titel: Liquid-based cervical cytologic smear study and conventional Papanicolaou smears: A metaanalysis of prospective studies comparing cytologic diagnosis and sample adequacy. Autoren: Bernstein SJ, Sanchez-Ramos L, and Ndubisi B Quelle: <i>Am J Obstet Gynecol</i> 2001; 185;308-317.				
Das vorliegende Dokument enthält: qualitative Informationssynthese X quantitative Informationssynthese X				
Klas	A. Fragestellung	Ja	Nein	?
QA	Ist die Forschungsfrage relevant für die eigene Fragestellung?	x		
Klas	B. Informationsgewinnung			
	Dokumentation der Literaturrecherche:	x		
QA	Wurden die genutzten Quellen dokumentiert?	x		
QB	Wurden die Suchstrategien dokumentiert?	x		
QB	Wurden Einschlusskriterien definiert?	x		
QB	Wurden Ausschlusskriterien definiert?	x		
Klas	C. Bewertung der Information			
	Dokumentation der Studienbewertung			
QA	Wurden Validitätskriterien berücksichtigt?		x	
QB	Wurde die Bewertung unabhängig von mehreren Personen durchgeführt?		x	
QC	Sind ausgeschlossene Studien mit ihrem Ausschlussgründen dokumentiert?		x	
QC	Ist die Datenextraktion nachvollziehbar dokumentiert?	x		
QC	Erfolgte die Extraktion von mehreren Personen unabhängig?		x	
Klas	D. Informationssynthese			
	Quantitative Informationssynthesen:	x		
QA	Wurde das Meta-Analyse Verfahren angegeben?	x		
QB	Wurden Heterogenitätstestungen durchgeführt?	x		
QC	Sind die Ergebnisse in einer Sensitivitätsanalyse auf Robustheit überprüft?		x	
	Qualitative Informationssynthesen:	x		
QA	Ist die Informationssynthese nachvollziehbar dokumentiert?	x		
QB	Gibt es eine Bewertung der bestehenden Evidenz?	x		
Klas	E. Schlussfolgerungen			
QB	Wird die Forschungsfrage beantwortet?	x		
QB	Wird die bestehende Evidenz in den Schlussfolgerungen konsequent umgesetzt?		x	
QA	Werden methodisch bedingte Limitationen der Aussagekraft kritisch diskutiert?	x		
I	Werden Handlungsempfehlungen ausgesprochen?		x	
I	Gibt es ein Grading der Empfehlungen?		x	
I	Wird weitere Forschungsbedarf identifiziert?	x		
I	Ist ein „Update“ des Review eingeplant?		x	
Klas	F. Übertragbarkeit der internationalen/ausländischen Ergebnisse und Schlussfolgerungen			
	Bestehen Unterschiede hinsichtlich der / des:			
	Epidemiologie der Zielkondition?		x	
	Entwicklungsstandes der Technologie?		x	
	Indikationsstellung?		x	
	Versorgungskontexte, -bedingungen, -prozess?	x		
	Vergütungssysteme?	x		
	Sozioökonomischen Konsequenzen?	x		
	Patienten- und Providerpräferenzen?			x
Beurteilung: Berücksichtigt:				

Titel: Assessment of Techniques for Cervical Cancer Screening.				
Autoren: Canadian Coordinating Office for Health Technology Assessment				
Quelle: CCOHTA Report 1997: 2E				
Dokumenttyp: Assessment Dokument				
Adressaten: Entscheidungsträger Kliniker X Patienten Andere				
Klas	A. Fragestellung und Kontext	Ja	Nein	?
I	Werden Anlass und Ziel der Publikation im Sinne einer „Policy Question“ dargestellt?	x		
QA	Gibt es im Rahmen des breiteren Kontext eine präzise formulierte Forschungsfrage nach der (interessierenden) Intervention?	x		
I	Sind in der Publikation Angaben zu folgenden Aspekten enthalten:			
I	Epidemiologie der Zielerkrankung	x		
I	(Entwicklungs-)stand der Technologie		x	
I	Efficacy	x		
I	Effectiveness	x		
I	Nebenwirkungen			
I	Indikationen			
I	Kontraindikationen			
I	Praxisvariation			
I	Versorgungsstrukturen	x		
I	Kostengesichtspunkten	x		
I	sozioökonomischen, ethischem und juristischen Impact	(x)		
Klas	B. Methodik der Informationsgewinnung			
QA	Wurden die genutzten Quellen dokumentiert?	x		
QB	Wurden die Suchstrategien dokumentiert?	x		
QB	Wurden Einschlusskriterien definiert?	x		
QA	Wurden Ausschlusskriterien definiert?	x		
Klas	C. Methodik der Bewertung der Dokumentation			
QA	Wurden Validitätskriterien berücksichtigt?		x	
QC	Wurde die Bewertung unabhängig von mehreren Personen durchgeführt?			x
QC	Sind ausgeschlossene Studien mit ihrem Ausschlussgründen dokumentiert?		x	
QC	Ist die Datenextraktion nachvollziehbar dokumentiert?		x	
QC	Erfolgte die Datenextraktion von mehreren Personen unabhängig?			x
Klas	D. Methodik der Informationssynthese			
I	Es wurden quantitative Informationssynthesen durchgeführt: (Metaanalyse-Bogen 1b)		x	
I	Es wurden qualitative Informationssynthesen durchgeführt: (Bogen 1b)	x		
I	Es wurden zur Ergänzung der Datenlage eigene Erhebungen durchgeführt:		x	
Klas	E. Ergebnisse/Schlussfolgerungen			
QB	Wird die bestehende Evidenz in den Schlussfolgerungen konsequent umgesetzt?	x		
QA	Werden methodisch bedingte Limitationen der Aussagekraft kritisch diskutiert?		x	
I	Werden Handlungsempfehlungen ausgesprochen?		x	
I	Gibt es ein Grading der Empfehlungen?		x	
QC	Wurde die Publikation vor der Veröffentlichung einem externen Reviewverfahren unterzogen?	x		
I	Ist ein „Update“ der Publikation eingeplant?		x	
Klas	F. Übertragbarkeit der internationalen/ausländischen Ergebnisse und Schlussfolgerungen			
	Bestehen Unterschiede hinsichtlich der / des:			
	Epidemiologie der Zielkondition?			
	Entwicklungsstandes der Technologie?			
	Indikationsstellung?			
	Versorgungskontexte, -bedingungen, -prozess?			
	Vergütungssysteme?			
	Sozioökonomischen Konsequenzen?			
	Patienten- und Providerpräferenzen?			
Beurteilung: Ausgeschlossen:				

II Ökonomische Evaluation

Verfasser:

Uwe Siebert^{1,2,3}, Gaby Sroczynski^{1,2}, Kerstin Voigt¹, Pamela Aidelsburger³, Jutta Engel⁴, Bernhard Gibis⁵, Peter Hillemanns⁶, Dieter Hölzel⁴, Jürgen Wasem³

- ¹ Bayerische Forschungs- und Koordinierungsstelle Public Health, Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Klinikum Großhadern, Ludwig-Maximilians-Universität, München
- ² Harvard Center for Risk Analysis, Harvard School of Public Health, Boston, USA
- ³ Institute for Technology Assessment, Massachusetts General Hospital, Harvard Medical School, Boston, USA
- ³ Alfried Krupp von Bohlen und Halbach-Stiftungslehrstuhl für Medizinmanagement, Universität Duisburg-Essen, Essen
- ⁴ Tumorzentrum München, Tumorregister, Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Klinikum Großhadern, Ludwig-Maximilians-Universität, München
- ⁵ Kassenärztliche Bundesvereinigung (KBV), Dezernat 2, Versorgungsqualität und Sicherstellung, Berlin
- ⁶ Klinik und Poliklinik für Frauenheilkunde und Geburtshilfe, Klinikum Großhadern, München

II C Hauptdokument Ökonomische Evaluation

II C 1 Policy Question

Die Policy Question für den gesamten HTA-Bericht wurde in Kapitel I C 1 beschrieben.

II C 2 Hintergrund

Der Hintergrund für den gesamten HTA-Bericht wurde detailliert in Kapitel I C 2 dargestellt.

II C 3 Forschungsfragen

Ziel der ökonomischen Evaluation ist es, die ökonomischen Konsequenzen des Einsatzes von Dünnschichtpräparationen und computergestützten Auswertungsverfahren zur Untersuchung von Zervixabstrichen im Rahmen des gesetzlichen Krebsfrüherkennungsprogramms anhand der derzeit verfügbaren wissenschaftlichen Evidenz zu beurteilen. Ergebnisse dieser Evaluation sollen zur gesundheitspolitischen Diskussion darüber, ob eine Aufnahme der Technologien in den Leistungskatalog der gesetzlichen Krankenversicherung sinnvoll erscheint, einen klärenden Beitrag liefern.

Im einzelnen sollen im Rahmen der ökonomischen Evaluation folgende Forschungsfragen beantwortet werden.

1. Welche Kosten werden derzeit in Deutschland für das Zervixkarzinom-Screening aufgewendet?
2. Wie hoch sind die Behandlungskosten des Zervixkarzinoms?
3. Wie ist die Datenlage für eine ökonomische Beurteilung der Dünnschichtpräparation und der automatisierten und halbautomatisierten Technologien der Zervixzellabstrich-Auswertung?
4. Was kostet die Durchführung der Dünnschichtpräparation und der automatisierten und halbautomatisierten Technologien?
5. Welche ökonomischen Konsequenzen in Bezug auf die Gesamtkosten hätte ein routinemäßiger Einsatz dieser Technologien in der derzeitigen gynäkologischen Krebsfrüherkennung in Deutschland?
6. Wie ist die Kosten-Effektivität der untersuchten Technologien zu bewerten?
7. In welchen Bereichen besteht weiterhin Forschungsbedarf?

II C 4 Methodik

II C 4. 1 Quellen und Recherchen

II C 4. 1. 1 Literaturdatenbanken

Folgende Datenbanken wurden im Rahmen der systematischen Literaturrecherche durchsucht: Medline, PreMedline, Pubmed, Embase, HealthStar, CINAHL, Cancerlit, DARE, NEED, EconLit, NHS HTA-Databases, Cochrane Library, Best Evidence, sowie Web of Science (Science Citation Index, Social Science Citation Index and Arts & Humanities Citation Index). Die genutzten Datenbanken mit den jeweils berücksichtigten Jahrgängen sind in Dokument 1 im Anhang aufgeführt.

Es wurde eine internationale Recherche und eine nationale Recherche nach deutschen bzw. deutschsprachigen Artikeln durchgeführt. Auf eine zeitliche Beschränkung der Suche wurde in der deutschsprachigen Recherche bewusst verzichtet. Die internationale Recherche beschränkte sich auf die Suche aktueller Publikationen, die in internationalen HTA-Berichten angegeben wurden und wurde vom Information Services Institute of Health Economics, Edmonton, Kanada durchgeführt. Es wurde eine Freitextsuche mit Schlagwörtern und deren Verknüpfungen mit AND/OR-Operatoren durchgeführt.

Die Suche erfolgte nach folgenden Themengruppen unterteilt:

A. Verfahren zur Früherkennung des Zervixkarzinoms:

- Klinische Effektivität konventioneller Papanicolaou-Tests, Dünnschichtpräparationsverfahren, der neuen computergestützten Auswertungsverfahren und Kombinationen dieser Technologien
- Auskünfte zur technischen Funktion neuer Technologien

B. Gesundheitsökonomie:

- Kosten der konventionellen Papanicolaou-Tests, Dünnschichtpräparationsverfahren, der neuen computergestützten Auswertungsverfahren und Kombinationen dieser Technologien zur Zervixkarzinom-Früherkennung.
- Ökonomische Evaluation der konventionellen Papanicolaou-Tests, Dünnschichtpräparationsverfahren, neue computergestützte Auswertungsverfahren und Kombinationen dieser Technologien zur Zervixkarzinom-Früherkennung.
- Kosten für das konventionelle Papanicolaou-Verfahren sowie für die Zervixkarzinomtherapie und -Nachfolgebehandlung in Deutschland.

C. Zervixkarzinom

- Epidemiologie des Zervixkarzinoms in Deutschland.
- Durchführung der Zervixkarzinom-Diagnostik und -Früherkennung, sowie -Therapie und -Nachbehandlung in Deutschland.
- Studien zur zytologisch-histologischen Progression bzw. Regression zervikaler präkanzeröser Läsionen sowie zur Erkrankungsprogression bei Zervixkarzinom-Entwicklung.

D. Entscheidungsanalytische Modelle:

- Modelle zur Effektivität und/oder Kosten-Effektivität der genannten Verfahren zur Zervixkarzinom-Früherkennung.
- Strukturell basierend auf dem natürlichen Krankheitsverlauf des Zervixkarzinoms.

Die genauen Suchstrategien zur Literaturrecherche in diesen Gruppen wurden in Dokument 1a im Anhang dokumentiert.

Literaturstellen der Gruppen A, C und D dienen als Hintergrundinformation zum Thema Zervixkarzinom und Screening, insbesondere für die Strukturentwicklung und Parameterbesetzung in der entscheidungsanalytischen Modellierung des Zervixkar-

zinom-Screenings. Diese Literaturstellen wurden keiner systematischen Bewertung unterzogen.

II C 4. 1. 2 HTA-Berichte und Reviews

Die in Tabelle 25 aufgeführten internationalen Einrichtungen wurden nach themenrelevanten HTA-Berichten und Reviews durchsucht.

Tabelle 25: Publikationsliste internationaler Einrichtungen

Internationale Einrichtung	Internetseite
Agency for Health Care Research and Quality (AHRQ)	http://www.ahrq.gov
Agencia de Evaluacion de Tecnologias Sanitarias (AETS)	http://www.isciii.es/aets
Basque Office for Health Technology Assessment, Health Department Basque Government (OSTEBA)	http://www.euskadi.net/sanidad
Canadian Coordinating Office for Health Technology Assessment (CCOHTA)	http://www.ccohta.ca
Cochrane Library	http://www.cochrane.org
CRD Health Technology Assessment Database (CRD)	http://www.york.ac.uk/inst/crd/
Deutsches Institut für Medizinische Dokumentation und Information (DIMDI)	http://www.dahta.dimdi.de
ECRI	http://www.ecri.org/
Healthcare Insurance Board / College voor zorgverzekeringen (CVZ)	http://www.cvz.nl
Health Services Utilization and Research Committee (HSURC)	http://www.sdh.sk.ca/hsurc/cervical.htm
HSTAT Health Services of the National Library of Medicine	http://text.nlm.nih.gov/
Minnesota Health Technology Advisory Committee (HTAC)	http://www.health.state.mn.us
National Coordinating Center for Health Technology Assessment (NCCHTA)	http://www.ncchta.org
National Horizon Scanning Center (NHSC)	http://www.hsrc.org.uk/horizon
National Institute for Clinical Excellence (NICE).	http://www.nice.org.uk
New Zealand Health Technology Assessment (NZHTA)	http://nzhta.chmeds.ac.nz
NHS Center for Reviews and Dissemination Databases (NHS)	http://www.hta.nhsweb.nhs.uk/

II C 4. 1. 3 Internetseiten, Kongressberichte und Richtlinien

Die in Tabelle 26 aufgeführten Internetseiten wurden nach relevanter (Hintergrund-) Information und epidemiologischen Daten, sowie Richtlinien und Empfehlungen durchsucht. Ferner wurden die Referenzlisten der Publikationen auf relevante Beiträge gesichtet.

Tabelle 26: Internetseiten der berücksichtigten Institutionen

Berücksichtigte Institution	Internetadresse
Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften	http://www.uni-düsseldorf.de/WWW/AWMF/
Copernic	http://www.copernic.com
Cancernet	http://cancernet.nci.nih.gov/
Cancerindex	http://www.cancerindex.org/
Gesundheitsberichterstattung Deutschland	http://www.gbe-bund.de
Oncolink	http://www.oncolink.upenn.edu/cancernet/
Robert-Koch-Institut	http://www.rki.de
Statistisches Bundesamt	http://www.StaBu.de
Saarländisches Krebsregister	http://www.krebsregister.saarland.de
Universität Jena	http://www.uni-jena.de/ufk/cd/folder/start.htm
Weltgesundheitsorganisation	http://www.who.int
Leitlinien Europa und Deutschland, deutsche Tumorregister, Zytologiefachgesellschaften	http://www.zytologie.de
Firma Accumed	http://www.accumed.com
Gynäkologische Praxis	http://www.zyto.de/ThinPrep.htm
Gynäkologische Praxis	http://www.helligrath.de
Gynäkologische Praxis	http://www.praxis-tauber.de
Zyto-Pathologisches Institut	http://www.patho.bonn.com
Firma Oncometrix	http://www.oncometrix.com
Firma Cytoc	http://www.cytoc.com
Firma TriPath Imaging	http://www.tripathimaging.com
Verband privater Krankenversicherer	http://www.pkv.de

II C 4. 2 Expertenkontakte

Epidemiologische Primärdaten wurden von den folgenden Experten zur Verfügung gestellt:

- Herr Prof. Dr. Dieter Hölzel, Tumorzentrum München, Tumorregister, Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Klinikum Großhadern, Ludwig-Maximilians-Universität, München
- Frau Dr. Jutta Engel, MPH, Tumorzentrum München, Tumorregister, Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Klinikum Großhadern, Ludwig-Maximilians-Universität, München
- Frau Dr. Bettina Eisinger, Leiterin der Registerstelle, Gemeinsames Krebsregister (GKR), Berlin.
- Herr Dr. Roland Stabenow, stellvertretender Leiter der Registerstelle, Gemeinsames Krebsregister (GKR), Berlin.
- Herr Dr. Hartwig Ziegler, Leiter der Registerstelle, Epidemiologisches Krebsregister Saarland, Saarbrücken.
- Frau Dr. Christa Stegmaier, stellvertretende Leiterin der Registerstelle, Epidemiologisches Krebsregister Saarland, Saarbrücken.
- Herr Prof. Dr. Hermann Brenner, MPH, Abteilung für Epidemiologie, Deutsches Zentrum für Altersforschung an der Universität Heidelberg (DZFA).

Als klinische Experten im Projekt eingebunden waren:

- Herr PD Dr. Peter Hillemanns, Klinik und Poliklinik für Frauenheilkunde und Geburtshilfe, Klinikum Großhadern, München.
- Herr Prof. Dr. Christian Thaler, Klinik und Poliklinik für Frauenheilkunde und Geburtshilfe, Klinikum Großhadern, München.

Die Erhebung der ökonomischen Primärdaten wurde von den folgenden Experten durchgeführt:

- Herr Prof. Dr. Jürgen Wasem, Lehrstuhl für Allgemeine Betriebswirtschaftslehre und Gesundheitsmanagement, Ernst-Moritz-Arndt-Universität Greifswald.
- Frau Dr. Pamela Aidelsburger, MPH, Lehrstuhl für Allgemeine Betriebswirtschaftslehre und Gesundheitsmanagement, Ernst-Moritz-Arndt-Universität Greifswald.

Die internationale Literaturrecherche wurde durchgeführt von:

- Frau Dr. Leigh-Ann Topfer, Director, Information Services Institute of Health Economics, Edmonton, Kanada.

In der entscheidungsanalytischen Modellierung erfolgte eine Zusammenarbeit mit folgenden Experten:

- Frau Prof. Dr. Sue J. Goldie, MPH, Harvard Center for Risk Analysis, Harvard School of Public Health, Boston, MA, USA.
- Herr Dr. Evan Myers, Department of Obstetrics and Gynecology, Center for Clinical Health Policy Research-Evidence-Based Practice Center and Division of General Internal Medicine, Duke University, Durham, NC, USA.

Zur Erfassung ökonomischer Daten der neuen Technologien in Deutschland wurden folgende Personen/Institutionen kontaktiert:

- Herr Hans Auer, Senior Product Manager, TriPath Imaging, Inc., Düsseldorf.
- Herr Helmut Goldbecker, Geschäftsführer, Medite Medizintechnik, Burgdorf.
- Herr Dr. Christoph Börsch, Arbeitsgemeinschaft zytologisch tätiger Ärzte, Frankfurt a. Main.

Darüber hinaus wurden folgende externe Experten kontaktiert und bezüglich unveröffentlichter Daten, Informationen und neuer Studien zum Thema befragt:

- Frau Dr. Stefanie Klug, MPH, Abteilung Epidemiologie und medizinische Statistik, Fakultät für Gesundheitswissenschaften, Universität Bielefeld.
- Frau Prof. Dr. Maria Blettner, Abteilung Epidemiologie und medizinische Statistik, Fakultät für Gesundheitswissenschaften, Universität Bielefeld.
- Frau Dr. Heike Hoyer, Institut für Medizinische Statistik, Informatik und Dokumentation, Universität Jena.
- Herr Prof. Dr. Ulrich Schenck, Zytologisches Institut der Bayerischen Krebsgesellschaft e.V.; Institut für Allgemeine Pathologie und Pathologische Anatomie der Technischen Universität München.
- Herr Dr. Georg Marckmann, MPH, Institut für Ethik und Geschichte der Medizin, Universität Tübingen.
- Frau Mirjam Brach, MPH, Klinikum der Universität München, Klinik und Poliklinik für Physikalische Medizin und Rehabilitation.

II C 4. 3 Systematische Literaturselektion

Die Selektion der im Zuge der Recherche identifizierten Literatur erfolgte in zwei Arbeitsschritten.

Zunächst wurden alle gefundenen Literaturstellen in eine Literaturdatenbank (EndNote[®] 5.0, ISI ResearchSoft[™]) importiert und von zwei Mitarbeitern des HTA-Teams durch Sichtung der Titel und Abstracts nach den in Tabelle 9 aufgeführten Ein- und Ausschlusskriterien vorselektiert. Im anschließenden zweiten Schritt wurde die Vorauswahl durch Sichtung der Volltextversionen weiter eingeschränkt.

Weiterhin wurden folgende Kriterien berücksichtigt: Auf Grund der ungenügenden Beschreibung der Methodik in Abstracts wurden ausschließlich Volltext-Publikationen eingeschlossen. Alle Publikationen sollten in deutscher oder englischer Sprache vorliegen und empirische Daten beinhalten.

Die aufgenommenen Literaturstellen wurden parallel von zwei Mitarbeitern des HTA-Teams unabhängig voneinander systematisch nach Titel und Abstract gesichtet. Bei

Diskrepanzen wurde ein dritter Gutachter hinzugezogen und ein Konsensusentscheid herbeigeführt.

Tabelle 27: Ein- und Ausschlusskriterien der Literatursuche. Gruppe A: Verfahren zur Früherkennung des Zervixkarzinoms

Einschlusskriterien

Es werden alle systematischen Reviews, wissenschaftlichen Publikationen, Peer-reviewed Publikationen, Internetinformationen sowie persönliche Auskünfte zur technischen Funktion der liquid-based und computergestützten Test- bzw. Auswertungsverfahren zytologischer Abstriche der Zervix eingeschlossen.

Klinische Studien, systematische Reviews und Metaanalysen, die Angaben zu Testcharakteristika, insbesondere über Sensitivität und Spezifität der Tests sowie zur Reliabilität und Validität der Messergebnisse geben, werden eingeschlossen.

Ausschlusskriterien

Publikationen, die sich auf andere Methoden als die konventionelle Papanicolaou-Zytologie, liquid-based Zytologie oder computergestützte Auswertungsverfahren beziehen bzw. sich auf den Einsatz der genannten Früherkennungsverfahren bei anderen Erkrankungen als dem Zervixkarzinom beziehen, werden ausgeschlossen. Vergleichsstudien zu verschiedenen Abstrichentnahmetechniken werden ausgeschlossen.

Alle Publikationen, die persönliche Bemerkungen ohne Bezug zu Originaldaten beinhalten, Briefe sowie reine Marketing- und Werbematerialien von Testherstellern, werden ausgeschlossen.

Alle Publikationen zu Testcharakteristika, die keine Angaben zur Studienpopulation machen, werden ausgeschlossen.

Alle Publikationen zu Testcharakteristika, die nicht Primär- oder Rescreening der genannten Verfahren evaluieren, werden ausgeschlossen.

Alle Publikationen zu Testcharakteristika, die keinen der aufgeführten Referenzstandards (Histologie, Kolposkopie oder Zytologie) verwendeten, werden ausgeschlossen.

Alle Publikationen zu Testcharakteristika, die keine Daten zu Sensitivität, Spezifität, prädikativen Werten, Reliabilität oder Validität liefern, werden ausgeschlossen.

Alle Publikationen zu Testcharakteristika, die den Einsatz der Screeningverfahren nach bereits therapiertem Zervixkarzinom bzw. nach Behandlung von präkanzerogenen Zellläsionen evaluieren, werden ausgeschlossen.

Alle Publikationen zur Evaluation der genannten Screeningverfahren, die sich auf Studienpopulationen mit HIV-Infektion bzw. unter Schwangerschaft beziehen, werden ausgeschlossen.

Tabelle 28: Ein- und Ausschlusskriterien der Literatursuche. Gruppe B: Gesundheitsökonomie

Einschlusskriterien

Eingeschlossen werden alle wissenschaftlichen Publikationen, die Daten zu direkten und indirekten Kosten der liquid-based Zytologie oder den computergestützten Auswertungsverfahren in der Zervixkarzinom-Früherkennung beinhalten.

Es werden alle Publikationen, die Daten zu direkten und indirekten Kosten der Zervixkarzinomtherapie und -nachfolgebehandlung in Deutschland sowie Kosten zur konventionellen Papanicolaou-Methode für Deutschland beinhalten, eingeschlossen.

Alle wissenschaftlichen Publikationen, die die ökonomische Evaluation der genannten neuen zytologischen Verfahren zur Zervixkarzinom-Früherkennung zum Gegenstand haben, werden berücksichtigt.

Systematische Reviews und Metaanalysen zu den genannten konventionellen und neuen Technologien werden eingeschlossen.

Ausschlusskriterien

Ökonomische Publikationen, die Kostendaten der Zervixkarzinom-Therapie und -Nachfolgebehandlungen in anderen Ländern als Deutschland berichten, werden ausgeschlossen.

Ökonomische Publikationen, die nicht aktuelle Kostendaten (vor 1990) der Zervixkarzinom-Therapie und -Nachfolgebehandlung berichten bzw. sich auf nicht den aktuellen Richtlinien entsprechenden Therapieformen in Deutschland beziehen, werden ausgeschlossen.

Ökonomische Publikationen, die sich auf andere Methoden als der konventionellen Papanicolaou-Zytologie (in Deutschland), liquid-based Zytologie oder den computergestützten Auswertungsverfahren beziehen bzw. unsystematische Reviews zu den genannten Verfahren sind, werden ausgeschlossen.

Alle Publikationen, die keine aktuellen Kostendaten (vor 1990) zu den genannten konventionellen und neuen Technologien angeben bzw. ökonomische Evaluationsstudien, die andere Technologien als die genannten evaluieren, werden ausgeschlossen.

Alle ökonomischen Publikationen, die sich auf Studienpopulationen mit HIV-Infektion bzw. unter Schwangerschaft beziehen, werden ausgeschlossen.

Alle ökonomischen Publikationen, die sich auf nicht übertragbare Gesundheitssysteme, insbesondere Entwicklungsländer beziehen, werden ausgeschlossen.

Ökonomische Publikationen, die sich auf den Einsatz der genannten Früherkennungsverfahren bei anderen Erkrankungen als dem Zervixkarzinom beziehen, werden ausgeschlossen.

Tabelle 29: Ein- und Ausschlusskriterien der Literatursuche. Gruppe C: Zervixkarzinom

Einschlusskriterien

Eingeschlossen werden dem aktuellen Wissensstand entsprechende deutsche Richtlinien und Empfehlungen zur Durchführung der Diagnostik und Therapie des Zervixkarzinoms.

Wissenschaftliche Publikationen und Reviews, die Angaben zur derzeitig praktizierten Zervixkarzinom-Diagnostik, sowie -therapie und -nachbehandlung in Deutschland beinhalten, werden eingeschlossen.

Alle wissenschaftlichen Publikationen, insbesondere klinische Studien, die Daten zur zytologisch-histologischen Progression bzw. Regression, Erkrankungsprogression des Zervixkarzinoms sowie deren Korrelation beinhalten, werden eingeschlossen.

Eingeschlossen werden alle wissenschaftlichen Publikationen, die aktuelle Angaben zur Prävalenz und Inzidenz des Zervixkarzinoms, der durch das Zervixkarzinom bedingten Mortalität sowie der altersabhängigen Distribution der Karzinomstadien in Deutschland enthalten.

Alle wissenschaftlichen Publikationen, die Angaben zur Prävalenz und Inzidenz zytologischer und histologischer Befunde präkanzerogener Vorstadien des Zervixkarzinoms in Deutschland enthalten, werden eingeschlossen.

Wissenschaftliche Publikationen mit Informationen zu Risikofaktoren und ihrer Verteilung, sowie zur Teilnehmerate an der Zervixkarzinom-Früherkennung in Deutschland werden eingeschlossen.

Ausschlusskriterien

Ausgeschlossen werden wissenschaftliche Publikationen, die nicht dem aktuellen Wissensstand entsprechende Angaben zur Durchführung von Diagnostik und Therapie des Zervixkarzinoms in Deutschland berichten oder Daten zu Diagnostik und Therapie des Zervixkarzinoms in anderen Ländern berichten.

Ausgeschlossen werden wissenschaftliche Publikationen, die nicht dem aktuellen Wissensstand entsprechende epidemiologische Daten zum Zervixkarzinom bzw. ihrer zytologisch-histologischen präkanzerogenen Vorstadien in Deutschland berichten oder Daten aus anderen Ländern berichten.

Ausgeschlossen werden alle wissenschaftlichen Publikationen zur Zervixkarzinom-Entstehung, die keine Daten zur zytologisch-histologischen Progression bzw. Regression und Erkrankungsprogression berichten.

Ausgeschlossen werden alle wissenschaftlichen Publikationen, die keine Daten zu den in den Einschlusskriterien beschriebenen Sachverhalten angeben.

Tabelle 30: Ein- und Ausschlusskriterien der Literatursuche. Gruppe D: Entscheidungsanalytische Modelle

Einschlusskriterien

Eingeschlossen werden alle wissenschaftlichen Publikationen, die entscheidungsanalytische Modelle zur Effektivität und/oder Kosten-Effektivität von Zervixkarzinom-Screeningverfahren beinhalten.

Eingeschlossen werden alle wissenschaftlichen Publikationen, die entscheidungsanalytische Modelle strukturell basierend auf dem natürlichen Krankheitsverlauf des Zervixkarzinoms, insbesondere Markov-Modelle, beinhalten.

Ausschlusskriterien

Ausgeschlossen werden alle wissenschaftlichen Publikationen mit entscheidungsanalytischen Modellen, die strukturell nicht auf dem natürlichen Krankheitsverlauf des Zervixkarzinoms begründet sind.

Ausgeschlossen werden alle wissenschaftlichen Publikationen zu mathematischen bzw. theoretischen Modellen, die keine entscheidungsanalytischen Modelle zur Effektivität und/oder Kosten-Effektivität von Zervixkarzinom-Screeningverfahren sind.

II C 4. 4 Bewertung der Literaturstellen

Berücksichtigte ökonomische Studien wurden in einem dritten Selektionsverfahren nach den folgenden Kriterien für die qualitative Beurteilung ausgewählt und anhand der im Anhang ersichtlichen Checkliste 3 zur Beurteilung der methodischen Qualität gesundheitsökonomischer Studien bewertet:

Systematische Reviews, HTA-Berichte und Primärstudien, die eine vollständige ökonomische Evaluations- bzw. Kostenstudie zu den liquid-based Technologien und computergestützten Auswertungsverfahren von Zervixabstrichen beinhalten, wie

- Kosten-Nutzen-Analysen,
- Kosten-Effektivitäts-Analysen und
- Kosten-Nutzwert-Analysen

wurden eingeschlossen.

Kommentare und Editorials wurden ausgeschlossen. Für eine vergleichende Bewertung gesundheitsökonomischer Studien wurden ausschließlich Kosten-Effektivitätsstudien, die die konventionelle Papanicolaou-Methode den genannten neuen Technologien gegenüberstellten, eingeschlossen. Hierbei war ein vergleichba-

rer Zeithorizont sowie die Angabe absoluter Werte für Kosten und Effekte oder die Angabe des inkrementellen Kosten-Effektivitäts-Verhältnisses (KEV) der neuen versus der konventionellen Technologie essentiell.

Auf die Literaturstellen der Gruppen A, C und D wurden keine Checklisten angewandt. Die Daten dieser Publikationen dienten der Hintergrundinformation zum Thema Zervixkarzinom und Screening sowie zur entscheidungsanalytischen Modellierung des Zervixkarzinom-Screenings.

Die Bewertung der inhaltlich-methodischen Qualität der Studien anhand der hierfür entwickelten Kriterienkataloge erfolgte zunächst durch zwei Projektmitarbeiter jeweils unabhängig voneinander. Im daran anschließenden Vergleich der Ergebnisse wurden Diskrepanzen im gemeinsamen Gespräch ausgeräumt. Konnte kein Konsens erzielt werden, wurde ein dritter Bewerter hinzugezogen.

Die für alle Katalogkriterien erhaltenen Punktwerte dienten als Basis zur Beurteilung der Studienqualität. Anhand einzelner Punktwerte konnte die Erfüllung einzelner Kriterien, anhand von Summenscores das Abschneiden in bewertungsrelevanten Teilbereichen sowie die Gesamtqualität beurteilt werden.

II C 4. 5 Datenextraktion und qualitative Synthese

Die wichtigsten Daten der ökonomischen Dokumente wurden systematisch extrahiert und im Sinne einer Datenbank zusammengestellt. Es wurde dafür die vorgegebene Datenbankstruktur des gesundheitsökonomischen Panels der German Scientific Working Group Technology Assessment for Health Care verwendet (Lühmann et al. 1999; Siebert U. et al. 1999).

Extrahierte Parameter eingeschlossener gesundheitsökonomischer Studien wurden in einer qualitativen Synthese entsprechend den gestellten Forschungsfragen als tabellarische Synopsis zusammengestellt.

Für die ökonomischen Evaluationen wurde die Synopsis auf folgende Merkmale verdichtet:

- Studie
- Land
- Währung
- Jahr
- Diskontrate
- Technologie
- Inkrementelle Effekte (in gewonnenen Lebenstagen)
- Inkrementelle Kosten (in US\$)
- Inkrementelles Kosten-Effektivitäts-Verhältnis (in US\$/Lebensjahr)

II C 4. 6 Währungskonversion und Inflationsbereinigung

Da die Ergebnisse zum Teil in unterschiedlichen Währungen vorlagen, wurden Währungskonversionen durchgeführt. Zur Umrechnung in US\$ wurden die von der OECD veröffentlichten Kaufpreisparitäten des Jahres 1996 für den Gesundheitssektor herangezogen (Quelle: OECD Health Data 2000). Auf zusätzliche Deflationierung wurde verzichtet, da die Kostenerhebungen aller Studien in einem Zeitraum von 2 Jahren stattfanden. Alle bereinigten Werte wurden sowohl für die durchschnittlichen als auch für die inkrementellen Maßzahlen angegeben. Die in der ökonomischen Primärdatenerhebung erfolgten Angaben in Deutsche Mark (DM) wurden mit dem Konversionsfaktor für Deutschland (1 Euro entsprechen 1,95583 DM) in Euro (€) umgerechnet.

II C 4. 7 Erhebung ökonomischer Primärdaten für Deutschland

Die Erhebung von ökonomischen Primärdaten wie Mengen- und Preisgerüste und ihre Zusammenführung zu Kostendaten für das Zervixkarzinom-Screening und die Zervixkarzinomtherapie in Deutschland wurde von den ökonomischen Experten der HTA-Arbeitsgruppe (Dr. P. Aidelsburger, Prof. Dr. J. Wasem) durchgeführt. Als

Kostenarten wurden ausschließlich direkte medizinische Kosten berücksichtigt. Aus Zeitgründen wurde innerhalb des Gesamtprojektes auf das Erfassen von indirekten Kostendaten verzichtet. Quellen für die in der Bestimmung der direkten Kosten zugrundeliegenden Mengen waren das Tumormanual „Empfehlungen zur Diagnostik, Therapie und Nachsorge – Zervixkarzinom“ (Tumorzentrum München 1999) und durch Expertenmeinung (Dr. Hillemanns, Universitätsklinikum München) ergänzt. Die Bewertung der Mengen erfolgte durch modifizierte administrative Preise als Substitute für Marktpreise.

Im stationären Bereich wurden die Verweildauern bewertet mit Pflegesätzen, ergänzt durch die Sonderentgelte (bei Hysterektomien) für Versicherte der gesetzlichen (GKV) und privaten Krankenversicherung (PKV) sowie die Abrechnung von Chefarztleistungen für den Anteil PKV-versicherter Frauen. Bei der ambulanten ärztlichen Versorgung wurden, entsprechend der Empfehlungen der Arbeitsgemeinschaft „Reha-Ökonomie“ (Hessel et al. 1999) die ärztlichen Leistungen mit einem Durchschnittspreis basierend auf der Gebührenordnung für Ärzte (GOÄ) und dem Einheitlichen Bewertungsmaßstab (EBM) bewertet. Der Durchschnittspreis ergab sich unter Berücksichtigung des GKV/PKV-Versichertenverhältnisses. Dabei wurden den unterschiedlichen Preisen in alten und neuen Ländern und dem Verhältnis von in GKV und PKV versicherten Frauen (getrennt nach Rechtskreisen Ost und West) Rechnung getragen.

Basisjahr bei der Erhebung der Preise war das Jahr 2000. Seit 1. Juli 1999 ist infolge von Reformen der Leistungserstattung in der GKV keine Bestimmung der Kosten einzelner Laborleistungen mehr möglich. Die Kostenberechnungen für Laborleistungen beruhen demzufolge auf Daten von 1999 vor Reform. Die Daten zur Berechnung der Gewichtung für PKV-Versicherte und GKV-Versicherte stammen aus dem Mikrozensus des Jahres 1999. Der durchschnittliche Pflegesatz wurde dem PKV-BV Zahlenbericht 2000 (Verband privater Krankenversicherer, <http://www.pkv.de>) entnommen. Bei allen Analysen wurde die gesamtgesellschaftliche Perspektive eingenommen.

Bestimmung der Mengen

Quelle für die in der Bestimmung der direkten Kosten zugrundeliegenden Mengen war das Tumormanual „Empfehlungen zur Diagnostik, Therapie und Nachsorge – Zervixkarzinom“ (Tumorzentrum München 1999), ergänzt durch Expertenmeinung (Dr. Hillemanns, Universitätsklinikum München). Die ermittelten Mengen wurden in einer Kostenvariablendatei erfaßt. Es wurde angestrebt, eine in Deutschland übliche

Standardpraxis wiederzuspiegeln. Dabei konnten individuelle im Einzelfall sinnvolle Untersuchungen bzw. Therapien nicht berücksichtigt werden.

Berechnung der Preise für GKV-versicherte Patienten

Grundlage für die Preisberechnung waren die im Einheitlichen Bewertungsmaßstab (Stand 1. Juli 1999) festgelegten Punktzahlen, die nach einem Punktwert in Deutschen Pfennigen multipliziert wurden. Basierend auf den Angaben der Kassenärztlichen Vereinigungen Berlin, Bayern, Nordrhein, Sachsen-Anhalt, Mecklenburg-Vorpommern und Südbaden wurde der Punktwert für 1999 mit durchschnittlich 7,5 Pfennig ermittelt. Aufgrund der Budgetierung wurde für nicht erstattbare Leistungen ein Abschlag von 12,7% (Quelle: Kassenärztliche Vereinigung Berlin) abgezogen. Der zur Berechnung der Preise verwendete Punktwert lag demnach bei 6,55 Pfennigen. In den Berechnungen für Beratung wurde ein Anteil von Rentnern mit 33,3% angenommen. Die Berechnung der Preise für Laborleistungen erfolgte anhand der EBM-Punktzahlen je Einzelleistung des bis 30.6.1999 geltenden EBM, da seit dem 1.7.1999 eine Einzelabrechnung von Laborleistungen nicht mehr möglich ist.

Berechnung der Preise für PKV-versicherte Patienten

Grundlage für die Preisberechnung waren die in der Gebührenordnung für Ärzte (Stand 1. Januar 2000) festgelegten Punktzahlen, die nach einem Punktwert in Deutschen Pfennigen multipliziert wurden. Die Punktwerte wurden nach dem Anteil der PKV-versicherten Frauen in alten und neuen Bundesländern differenziert verwendet. Steigerungsfaktor mit dem das Produkt aus Punktzahl und Punktwert multipliziert werden kann war der Regelhöchstsatz, also das 2,3-fache der einfachen Punktzahl.

Arzneimittelpreise

Für apothekenvertriebene Arzneimittel wurden die Apothekenverkaufspreise der Roten Liste 2001 für die größte frei verkäufliche Packungsgröße verwendet. Von diesen wurden 5% Krankenkassenrabatt für den Anteil der GKV-Versicherten Patientinnen abgezogen.

Bestimmung der Kosten der neuen Screeningtechnologien

Für computergestützte Auswertungsverfahren und Dünnschichtpräparationsverfahren wurde eine detaillierte Kostenaufstellung (microcosting) durchgeführt. Zu diesem Zweck erfolgte eine Kontaktaufnahme mit der Firma „Medite Medizintechnik“ (Burgdorf), sowie der Arbeitsgemeinschaft zytologisch tätiger Ärzte in Deutschland e.V. (AZÄD; Frankfurt a. M.). Die Firma „Medite Medizintechnik“ vertreibt das Dünnschichtpräparationsverfahren AutoCytePrep System und das computergestützte Auswertungsverfahren AutoPap GS Screening System. Die Arbeitsgemeinschaft zytologisch tätiger Ärzte gründete die Cytomed GbR, ein Zusammenschluss von ca. 30 zytologischen Laboren, die gemeinschaftlich die Präparationsleistung erbringt. In dieser wird das Dünnschichtpräparationsverfahren ThinPrep™ verwendet. Sowohl von „Medite Medizintechnik“ als auch von der Arbeitsgemeinschaft zytologisch tätiger Ärzte in Deutschland e.V. wurden Kostendaten zur Verfügung gestellt, die Grundlage der Kostenberechnung waren.

Die Kostendaten umfassten Anschaffungskosten, Materialkosten, Instandhaltungskosten, Personal- und Raumkosten. Eine Berechnung der Kosten pro Präparat erfolgte mit 3% jährlicher Verzinsung und 3% jährlicher Diskontierung von Kapitalkosten und unter Berücksichtigung von Abschreibungen für gekaufte Geräte, sowie 3% jährlicher Diskontierung von Instandhaltungskosten.

II C 5 Ergebnisse

II C 5. 1 Ergebnisse der systematischen Literatursichtung

II C 5. 1. 1 Ergebnisse der Literaturselektion

Insgesamt wurden nach systematischer Selektion 356 Literaturstellen, die die beschriebenen Einschlusskriterien erfüllten, berücksichtigt. In Tabelle 31 ist eine Übersicht der Anzahl aufgenommener Literaturstellen nach Themen-Gruppen unterteilt angegeben. Insbesondere HTA-Berichte treten in den jeweiligen Themengruppen mehrfach auf.

Tabelle 31: Anzahl, Publikationsart und Gruppenzugehörigkeit eingeschlossener Literatur

Gruppe	Literaturstellen insgesamt	Zeitschriftenartikel	HTA-Berichte, systematische Reviews	Buchartikel
A	242	230	10	2
B	30	21	9	0
C	67	65	2	0
D	36	31	5	0

Nicht alle Publikationen konnten im Volltext beschafft werden, ein HTA-Bericht (Hyde et al. 1999) ist noch nicht veröffentlicht worden.

Zur Themengruppe A (Screeningverfahren) wurden insgesamt 242 Publikationen gefunden, von denen 210 in der Volltextversion beschafft werden konnten: (Anonymous 1998; Abulafia & Sherer 1999a; Abulafia & Sherer 1999b; ACOG 1998; AHTAC 1998; Anderson T.L. & Nelson 1994; Anderson G. et al. 1997; Anonymous 1992; Anonymous 1996; Aponte-Cipriani et al. 1995; Ashfaq R & Saliger 1993; Ashfaq R. et al. 1995; Ashfaq R. et al. 1997; Ashfaq R. et al. 1999; Austin 1998; Austin & Ramzy 1998; Awen et al. 1994; Bai et al. 2000; Baird 1997; Bajardi 1979; Baldauf et al. 1995; Bartels P. H. et al. 1998; Bartels PH 1999; Bartels P. H. 2000; Bartoo et al. 1992; Bayrle 1977; Bednov et al. 1998; Bedrossian 1998; Beier & Kranzfelder 1989; Bergeron et al. 2000; Bibbo et al. 1999; Bishop 1997; Bishop et al. 1998;

Bishop et al. 1999a; Bishop et al. 1999b; Bishop et al. 2000; Bolick & Hellman 1998; Boon & Kok 1993; Boon et al. 1994; Boon et al. 1995; Bosanquet et al. 1999; Braun 1999; Broadstock 2000; Bur et al. 1995; Canavan & Doshi 2000; Carpenter & Davey 1999; Carter S.L. 1994; Cenci et al. 1997; Cenci et al. 2000; Cenci & Vecchione 2000; Coleman 1998; Chevront et al. 1998; Colgan et al. 1995; Cuzick et al. 1999; Cuzick et al. 2000; Diaz-Rosario & Kabawat 1999; DiBonito et al. 1993; Doornewaard et al. 1997a; Doornewaard et al. 1997b; Doornewaard et al. 1999; Doornewaard et al. 2000; Duggan & Brasher 1997; Duggan 2000; Dunton 2000; Dupree et al. 1998; ECRI 1999; ECRI 2000; Fahey et al. 1995; Farnsworth et al. 1996; Ferenczy et al. 1996; Ferris et al. 2000; Fetterman et al. 1999; Gay et al. 1985; Geyer et al. 1993; Ghidoni et al. 1998; Gornall et al. 2000; Grussendorf Conen 1993; Guidos & Selvaggi 1999; Guidos & Selvaggi 2000; Gullotta et al. 1997; Halford J.A. et al. 1997; Halford 1998; Halford J.A. et al. 1999; Hoerl et al. 2000; Horn et al. 1995; Howell et al. 1998; Howell et al. 1999; HTAC 1999; Huang et al. 1999; Hutchinson et al. 1991; Hutchinson et al. 1992; Hutchinson et al. 1995; Hutchinson et al. 1999; James & Knapp 1992; Jenny et al. 1997; Johnson et al. 1993; Jones & Novis 1996; Kaminsky et al. 1995; Kaser et al. 1981; Kesic et al. 1993; Kestler et al. 1999; Keyhani-Rofagha et al. 1996; Khalbuss et al. 2000; Kharazi et al. 1993; Knowles et al. 1992; Kok & Boon 1996; Kok et al. 1998; Koroschetz 1991; Koss L.G. et al. 1994; Koss L.G. et al. 1997; Koss 2000; Kraus & Schneider 1984; Kreuger et al. 1999; Kunz et al. 1998; Kurtycz & Hoerl 2000; Laverty et al. 1995; Laverty et al. 1997; Lee K.R. et al. 1997; Lee J.S. et al. 1998; Lerma et al. 1998; Lieu 1996; Linder 1997; Linder et al. 1997; Linder & Zahniser 1997; Linder 1998; Linder & Zahniser 1998; Lonky et al. 1999; Mango 1994; Mango 1996a; Mango 1996b; Mango 1997; Mango & Radensky 1998; Mango & Valente 1998; Marshall et al. 1999; Mayeaux Jr et al. 1995; McCrory & Matchar 1999; McGoogan & Reith 1996; McGoogan et al. 1998; Melamed & Kamentsky 1969; Melnikow J. et al. 1997; Melnikow J. & Nuovo 1999; Michelow et al. 1997; Minge et al. 2000; Mitchell & Medley 1998; Montz et al. 1992; Nagy & Keller 1989; Nanda et al. 2000; Neumann H. G. & Seidenschnur 1987; Neumann H. G. et al. 1989; NHS 1999; Noorani et al. 1997; Nyirjesy et al. 1998; O'Leary et al. 1998; Ouwerkerk-Noordam et al. 1994; Papillo et al. 1998; Patten S. et al. 1996a; Patten S.F. et al. 1996b; Patten S.F. et al. 1997; Payne et al. 2000; Perlman 2000; Pete et al. 1998; Quantin et al. 1992; Quitllet et al. 1997; Richart et al. 1996; Roberts et al. 1997; Roberts et al. 1999; Rodriguez Castro & Garcia Gonzales 1978; Ronco et al. 1996; Rosenthal et al. 1996; Rosenthal 1998; Rummel et al. 1977; Sasieni 1999; Sawaya & Grimes 1999; Sawaya & Washington 1999; Sawaya et al. 2001; Schenck 2000; Schenck & von Karsa 2000; Schlegel & Kuhndel 1983; Schneider A. & Zahm 1996; Schneider A. et al. 1996; Sheets et al. 1995; Sherman et al. 1994; Sherman et al. 1997; Shield & Cox 1998; Sigurdsson 1995; Slawson et al. 1994; Solomon H.M. & Frist 1998; Solomon D. et al. 2001; Soost H. J. et al. 1989; Soost H.J. et al. 1991; Spitzer 1998; Sprenger 1996; Steps 1982; Stevens M.W. et al. 1997a; Stoler 2000; Stoler & Schiffman 2001; Tawa et al. 1988; Troni et al. 2000; Van Duyne et al. 1989; Vassilakos et al. 1999; Vassila-

kos et al. 2000; Weintraub & Morabia 2000; Wertlake 1999; Widra et al. 1994; Wilbur D.C. et al. 1996b; Wilbur D. et al. 1996a; Wilbur D.C. et al. 1997; Wilbur D.C. et al. 1998; Wilbur D.C. et al. 1999; Wunder & Quaeitzsch 1973).

In Gruppe B (Kosten und ökonomische Evaluationen) erfüllten insgesamt 30 Publikationen die Einschlusskriterien, von denen 22 im Volltext erhältlich waren: (AHTAC 1998; Broadstock 2000; Brotzman et al. 1999; Brown & Garber 1999; ECRI 1998a; ECRI 1998b; ECRI 1999; ECRI 2000; Hutchinson et al. 2000; McCrory & Matchar 1999; McQuarrie et al. 2000; Meerding et al. 2001; Myers et al. 2000b; Noorani et al. 1997; O'Leary et al. 1998; Payne et al. 2000; Raab et al. 1999b; Radensky & Mango 1998; Schechter 1996; Smith et al. 1999; Troni et al. 2000; Montz et al. 2001).

Nach dem weiteren manuellen Selektionsschritt wurden von diesen 22 Studien 11 Publikationen (AHTAC 1998; Brown & Garber 1999; Hutchinson et al. 2000; McCrory & Matchar 1999; Noorani et al. 1997; Payne et al. 2000; Raab et al. 1999b; Radensky & Mango 1998; Schechter 1996; Smith et al. 1999; Montz et al. 2001) selektiert, die den a priori vorgegebenen Einschlusskriterien für eine Bewertung anhand der im Anhang ersichtlichen Checkliste zur methodischen Qualität gesundheitsökonomischer Studien entsprachen (Tabelle 34).

Ausgeschlossen wurden nach den beschriebenen Kriterien 11 Berichte und zugehörige Kontextdokumente (Broadstock 2000; Brotzman et al. 1999; Brown & Garber 1998; Check 1997; ECRI 1998a; ECRI 1998b; ECRI 1999; ECRI 2000; Hyde et al. 1999; Kisner 1994; McQuarrie et al. 2000; Meerding et al. 2001; Myers et al. 2000b; O'Leary et al. 1998; Pallatroni 1999; Troni et al. 2000). Die Ausschlussgründe sind in Tabelle 32 angegeben.

Tabelle 32: Ausgeschlossene ökonomische Studien mit Ausschlussgrund

Studie	Ausschlussgrund
Broadstock 2000	Keine vollständige Kostenstudie; reiner systematischer Review ohne eigene Berechnungen
Brotzman et al. 1999	Keine vollständige ökonomische Evaluations- bzw. Kostenstudie
Brown & Garber 1998	Nicht erhalten
Check 1997	Keine Kostendaten enthalten
ECRI 1998 a,b	Keine vollständige ökonomische Evaluations- bzw. Kostenstudie
ECRI 2000 a,b	Keine vollständige ökonomische Evaluations- bzw. Kostenstudie
Hyde et al. 1999	Keine vollständige ökonomische Evaluations- bzw. Kostenstudie
Kisner 1994	Keine vollständige ökonomische Evaluations- bzw. Kostenstudie
McQuarrie et al. 2000	Keine vollständige ökonomische Evaluations- bzw. Kostenstudie
Meerding et al. 2001	Keine vollständige ökonomische Evaluations- bzw. Kostenstudie
Myers 2000b	Keine Kostendaten, reiner Artikel zur entscheidungsanalytischen Modellierung der Zervixkarzinomkrankung
O'Leary et al. 1998	Keine vollständige ökonomische Evaluations- bzw. Kostenstudie
Pallatroni et al. 1999	Keine vollständige ökonomische Evaluations- bzw. Kostenstudie
Troni et al. 2000	Keine vollständige ökonomische Evaluations- bzw. Kostenstudie

Im Anhang (Tabelle 47) sind die wichtigsten Daten der eingeschlossenen gesundheitsökonomischen Publikationen dokumentiert.

Anhand der Einschlusskriterien der Gruppe C (Zervixkarzinom) wurden 65 Publikationen einbezogen. Im Volltext lagen 52 Veröffentlichungen vor (Anderl & Ledermair 1982; Arends et al. 1998; Baltzer 1998; Bappert 1982; Beckmann et al. 2000; Berndt & Hendel 1984; Bierfreund 1986; Blohmer 1999; Brenner et al. 1990a; Brenner et al. 1990b; Brenner et al. 1993; Burghardt et al. 1994; Burghardt & Östör 1983; Dietl et al. 1981; Ebeling 1983; Ebeling 1986; Ebeling & Sarembe 1986; Ebeling 1987; Ebeling & Nischan 1987; Ebeling et al. 1987; Foreman & Hader 1997; Gatta et al. 1999; Gitsch et al. 1993; Gustafsson et al. 1997; Heinzl 1989; Heinzl & Hirsbrunner 1991; Hillemanns et al. 1997; Hillemanns et al. 2000; Hochuli & Ragoth 1990; Hofmann et al. 1989; Kainz et al. 1995; Kavungu et al. 1983; Kolles et al. 1989; Kreienberg & Mobus 1995; Küppers & Schnurch 1997; Lasnik & Breitenecker 1987; Ludwig 1989; Meulman et al. 1992; Munck 1985; Neumann G. 1988; Östör 1993; Paterok et al. 1995; Petry et al. 1996; Promer et al. 1993; Quaas & Heinrich 1998; Robra & Dierks 1990; Schneider A. & Shah 1989; Schneider A. et al. 1999; Schneider & al. 2001; Sommer et al. 1992; Stoll 1989; Zimmer et al. 1987).

In Gruppe D (Internationale entscheidungsanalytische Modelle) wurden 36 Publikationen eingeschlossen, 31 Publikationen lagen im Volltext vor (Albert 1981; Brown & Garber 1999; Carter P.M. et al. 1993; Cuzick et al. 1999; Eddy 1987; Eddy 1990; Fahs et al. 1992; Goldie S.J. et al. 1999; Goldie S. et al. 2001; Johnson et al. 1993; Koopmanschap 1990; Mandelblatt J.S. & Fahs 1988; Mandelblatt J. et al. 1996; Matsunaga 1997; McCrory & Matchar 1999; McQuarrie et al. 2000; Melnikow J & Nuovo 1996; Myers et al. 2000a; Myers et al. 2000b; Raab 1997; Raab et al. 1999a; Radensky & Mango 1998; Sato et al. 1999; Schechter 1996; Smith et al. 1999; Taylor et al. 2000; van Ballegooijen et al. 2000; Van den Akker-van Marle et al. 1999; Wagner 1997; Wilson & Woodman 1995).

Tabelle 33: Manuelle Selektion zur qualitativen Bewertung ökonomischer Studien (Gruppe B)

	Literaturstellen insgesamt	HTA-Berichte, systematische Reviews	Zeitschriftenartikel
Vorselektion	22	7	15
Nachselektion	11	4	7

II C 5. 1. 2 Literaturbewertung

Eine detaillierte Übersicht der Einzelergebnisse der Studien-Qualitätsbewertung wird in Anhang II in Tabelle 48 gegeben. In den Spalten wurden die Punktwerte für jedes der 56 Bewertungskriterien (siehe Qualitätskatalog) studienweise aufgelistet. Zeilenweise wurde der durchschnittliche prozentuale Erfüllungsgrad der einzelnen Bewertungskriterien (% der maximalen Punktzahl) ermittelt. Bei allen Prozentberechnungen wurden nicht relevante Punkte aus dem Nenner ausgeschlossen.

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

- Bewertung der Gesundheitszustände mit Präferenzen (Kriterium 19)
- Durchführung von Währungskonversionen (Kriterium 29).

Unvollständig erfüllt (<40%) wurden die Qualitätskriterien bezüglich

- Durchführung von multivariaten Sensitivitätsanalysen (Kriterien 43)
- Einsetzen adäquater statistischer Inferenzmethoden für stochastische Daten (Kriterium 47)
- Diskussion von Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis (Kriterium 50)
- Diskussion wichtiger ethischer Verteilungsfragen (Kriterium 53).

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

- Präzise Formulierung der Fragestellung (Kriterium 1)
- Untersuchung sowohl von Kosten als auch von Gesundheitseffekten (Kriterium 9)
- Angabe einer sinnvollen Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt (Kriterium 38)
- Durchführung univariater Sensitivitätsanalysen (Kriterium 42).

Tabelle 34 aggregiert die Einzelitems der 11 im Qualitätskatalog festgelegten Bewertungsdimensionen und fasst diese studienweise (Spalten) zu einem prozentualen Qualitätsscore zusammen. Dabei werden alle Bewertungsdimensionen, ungeachtet der Anzahl ihrer Einzelpunkte, gleich gewichtet. Zusätzlich wurde der prozentuale Erfüllungsgrad der einzelnen Bewertungsdimensionen im Mittel aller Studien ermittelt (Zeilen).

Tabelle 34: Prozentuale Erfüllung der Qualitätskriterien und Summenscore evaluierter ökonomischer Studien

Qualitätskriterien	Schechter 1996	Noorani et al. 1997	AHTAC 1998	Radensky et al. 1998	Brown & Garber 1999	McCrory et al. 1999	Raab et al. 1999	Smith et al. 1999	Hutchinson et al. 2000	Payne et al. 2000	Montz et al. 2001	Mittelwert (% der max. Punktzahl)
Fragestellung	100%	75%	100%	75%	100%	100%	100%	100%	100%	100%	100%	95%
Evaluationsrahmen	100%	81%	69%	63%	100%	100%	56%	100%	100%	100%	88%	87%
Analysemethoden und Modellierung	88%	63%	63%	88%	88%	100%	88%	100%	75%	88%	75%	83%
Gesundheitseffekte	60%	50%	50%	70%	50%	80%	60%	60%	70%	80%	70%	64%
Kosten	71%	50%	13%	57%	86%	93%	57%	50%	50%	64%	64%	60%
Diskontierung	88%	0%	33%	100%	100%	100%	50%	63%	100%	75%	100%	73%
Ergebnispräsentation	71%	71%	71%	71%	93%	100%	86%	71%	86%	100%	71%	81%
Behandlung von Unsicherheiten	90%	42%	70%	50%	80%	120%	70%	60%	50%	80%	50%	69%
Diskussion	43%	29%	57%	36%	57%	93%	36%	36%	29%	64%	50%	48%
Schlussfolgerungen	75%	100%	100%	50%	100%	100%	100%	100%	75%	100%	100%	91%
Qualitätsscore	75%	55%	56%	63%	82%	96%	63%	68%	69%	82%	71%	71%
Rang	3	9	8	7	2	1	7	6	5	2	4	

Der durchschnittliche Qualitätsscore beträgt 71% mit einer Standardabweichung von 15%, wobei sich eine individuelle Variationsbreite von 55% bis 96% feststellen lässt. Den niedrigsten Gesamtscore erzielte dabei die Studie von Noorani et al. (1997) mit 55% und die Studie von AHTAC (1998) mit 56%. Der Gesamtscore der hinsichtlich der methodisch-inhaltlichen Qualität am besten abschneidenden Untersuchung von McCrory et al. (1999) betrug 96%, gefolgt von Payne et al. (2000) und von Brown & Garber (1999) mit einem Qualitätsscore von jeweils 82%.

Die Variationsbreite des durchschnittlichen Erfüllungsgrades der Qualitätsdimensionen reicht von 48% bis 95%. Deutliche Schwächen hinsichtlich der methodisch-inhaltlichen Qualität lassen sich im Durchschnitt in folgenden Bereichen erkennen (in Klammern der durchschnittliche prozentuale Erfüllungsgrad):

- Diskussion (48%) - die Variationsbreite über die Studien hinweg liegt zwischen 29% und 93%
- Kosten (60%) - die Variationsbreite über die Studien hinweg liegt zwischen 13% und 93%
- Gesundheitseffekte (64%) - die Variationsbreite über die Studien hinweg liegt zwischen 50% und 80%

Eine hohe methodisch-inhaltliche Qualität findet sich dagegen in den Bereichen:

- Fragestellung (95%) - die Variationsbreite über die Studien hinweg liegt zwischen 75% und 100%
- Schlussfolgerung (91%) - die Variationsbreite über die Studien hinweg liegt zwischen 50% und 100%

II C 5. 2 Kosten der Zervixkarzinom-Früherkennung und Therapie in Deutschland

In der Literatur wurden keine Publikationen zu aktuellen Kosten im Bereich der Zervixkarzinom-Früherkennung und der Zervixkarzinom-Therapie in Deutschland identifiziert.

Die von den ökonomischen Experten der HTA-Arbeitsgruppe durchgeführte Erhebung der ökonomischen Primärdaten ergaben die in Tabelle 35 aufgeführten middle-

ren Gesamtkosten für die Zervixkarzinom-Früherkennungsuntersuchung sowie für die Behandlung inklusive prätherapeutischer Diagnostik, Medikation und Nachsorge bei Auftreten eines Zervixkarzinoms. Mit Ausnahme der Medikations- und Nachsorgekosten, die jährlich anfallen, sind alle dargestellten Kosten als einmalige Kosten bei Leistungserbringung anzusehen.

Für eine Zervixkarzinom-Früherkennungsuntersuchung bestehend aus der gynäkologischen Früherkennungsuntersuchung mit zytologischer Abstrichentnahme und der zytologischen Auswertung werden derzeit in Deutschland basierend auf der Gebührenverordnung für Ärzte bzw. des Einheitsbewertungsmaßstabs im Mittel 17,28 € gezahlt. Die Kosten für eine Therapie bei Auftreten eines Karzinoms ist stadiumsabhängig und variiert bei aktuellen Therapierichtlinien und Empfehlungen. Es wurden mittlere Gesamtkosten zwischen 2000 € und 9051 € berechnet. Jährliche Medikationskosten umfassen lokale und systemische Östrogene, Analgetika sowie regulierende Medikamente nach Therapie der Karzinomstadien IB bis IV und variieren stadiumsabhängig zwischen 216 € bis 749 € jährlich. Die jährlichen Kosten der Nachsorgebehandlungen für Therapierte staffeln sich in Abhängigkeit von der Zeit nach erfolgter Therapie und liegen zwischen 2007 € in den ersten beiden Jahren und 49 € ab dem sechsten post-operativen Jahr.

Tabelle 35: *Mittlere Gesamtkosten für Früherkennungsuntersuchung, Therapie, Medikation und Nachsorge des Zervixkarzinoms in Deutschland (gerundete Werte)*

Leistung	Kosten (€)
Früherkennungsuntersuchung	
Gynäkologische Untersuchung im Rahmen der gesetzlichen Krebsfrüherkennung, Zytologischer Abstrich, zytologische Färbung und zytopathologische Bewertung	17
Therapie	
Therapie Stadium I A	2000
Therapie Stadium I B, II A, B	6382
Therapie Stadium III	9051
Therapie Stadium IV	7883
Medikation	
Medikation in Stadium I B, II A, B	216
Medikation in Stadium III	748
Medikation in Stadium IV	749
Nachsorge	
1. bis 2. Jahr p. o.	2007
3. bis 5. Jahr p. o.	503
Ab dem 6. Jahr p. o.	49

II C 5. 3 Kosten der neuen Technologien

Veröffentlichte Zahlen zu Kosten der genannten zu evaluierenden Technologien wurden nur in internationaler Literatur (Anonymous 1998; Brotzman et al. 1999; Brown & Garber 1999; ECRI 1999; ECRI 2000; Hutchinson 1996a; Hutchinson & Farber 2000; O'Leary et al. 1998) und auf Internetseiten internationaler Herstellerfirmen gefunden.

Diese Angaben variieren zum Teil beträchtlich (Tabelle 36). In internationalen Kosten-Effektivitäts-Studien wurden im Vergleich zum konventionellen Papanicolaou-Test inkrementelle Kosten im Bereich von 3–30 US\$ pro Test angegeben. Dabei umfassen die angegebenen Kosten die Material- und Durchführungskosten pro Zellabstrich. Es konnten keine Kostendaten für die Technologien AutoCyte SCREEN™ (TriPath Imaging), AcCell Series 2000™ (Ampersand Medical) und Pathfinder System™ (CompuCyte Corp.) aus der Literatur extrahiert werden.

Tabelle 36: Anbieter und Preisangaben neuer Technologien.

Handelsname	Firma	Inkrementelle Kosten [US\$]	Quelle
Dünnschichtpräparate			
ThinPrep	CYTYC Corp. Boxborough, MA, USA	9,75 – 30,00	(Hutchinson et al. 2000; ECRI 1999; Brown & Garber 1999)
AutoCyte PREP	TriPath Imaging Burlington, NC, USA	7,95	(ECRI 1999)
Computergestützte Bildanalyseverfahren			
PAPNET	Neuromedical Systems, Inc. Suffern, NY, USA	9,38 – 18,00	(Hutchinson 1996b; O'Leary et al. 1998; Brown & Garber 1999; Brotzman et al. 1999)
AutoPap	TriPath Imaging Burlington, NC, USA	3,00 – 14,75	(Brown & Garber 1999; Hutchinson 1996b; Hutchinson et al. 2000; ECRI 2000)

In einem englischen HTA-Bericht von Payne et al. (Payne et al. 2000) werden die in Tabelle 37 aufgeführten Kosten für die Dünnschichtpräparationen basierend auf Kosten für Verbrauchsmaterial und Kapitalkosten berichtet.

Tabelle 37: Kosten der Dünnschichtpräparationen nach Payne et al. 2000 (Payne et al. 2000)

	ThinPrep 2000	AutoCytePrep
Jahreskapazität (Präparate)	50000	60000
Lebensdauer (Jahre)	6	6
	Kosten (€)	Kosten (€)
Kapitalkosten	30000,00	45000,00
Jahreskosten Unterhaltung	-	3000,00
Kapitalkosten pro Einheit	0,10	0,13
Verbrauchsmaterial	3,50	3,00
Kosten pro Untersuchung	3,60	3,13

In der deutschen Literatur konnten keine Kostendaten zu den neuen Technologien gefunden werden. Derzeit ist eine Bewertungsziffer im Einheitlichen Bewertungsmaßstab (EBM) weder für die Dünnschicht-Präparationsverfahren noch für die halb- und vollautomatischen Zervixzellabstrich-Auswertungstechnologien gegeben, da diese Verfahren bislang keine regulären Kassenleistungen sind.

Versicherte der GKV müssen für eine Dünnschichtpräparation und eine automatisierte Auswertung einen Betrag von umgerechnet bis zu 51 € selbst entrichten

(Anonymous 1998) Die Kosten für Privatversicherte betragen umgerechnet 77 € (<http://www.tripathimaging.com>) und können im Einzelfall erstattet werden.

Erhebungen ökonomischer Primärdaten für die in Deutschland präsenten Technologien

Eine Berechnung der Kosten pro Präparat erfolgte für die in Deutschland vertretenen Technologien ThinPrep™, AutoCytePrep und AutoPap GS Screening System auf Grundlage der erhaltenen Kostendaten von der Arbeitsgemeinschaft zytologisch tätiger Ärzte in Deutschland e.V. (AZÄD; Frankfurt a. M), sowie der Firma Medite Medizintechnik (Burgdorf).

Die Kostenberechnungen der verschiedenen Technologien basieren auf unterschiedlichen Konzepten und unterschiedlich detaillierter zur Verfügung stehender Datenlage. Ein direkter Vergleich der verschiedenen Technologien ist aus diesem Grund nur schwer möglich.

Für die ThinPrep™-Technologie wurden Kosten für den Kauf eines Prozessors von 62.870 € Brutto (Stand September 2001) und eine jährliche Untersuchungskapazität von 40.000 Präparaten angegeben. Die Kosten des Verbrauchsmaterials bestehend aus einem Behältnis zur Aufnahme des Exfoliats, einem speziellen Entnahmegerät und dem Mikroporenfilter werden mit 7,55 € Brutto angegeben. Die in Tabelle 38 angegebene Kostenberechnung basiert auf den Angaben von CytoMed GbR, einem Zusammenschluss von ca. 30 zytologischen Laboren, die gemeinschaftlich die Präparationsleistung erbringen. Die Kostenberechnung geht von einer kostenlosen Bereitstellung eines Geräteprozessors bei einer Abnahmepflicht von Verbrauchsmaterial für 18.000 Präparationen innerhalb von 3 Jahren aus. Die Anschaffungskosten des Prozessors wurden auf das Material anteilmäßig umgelegt. In den Berechnungen sind die Kosten für die zytologische Auswertung der Präparate nicht enthalten.

Tabelle 38: Kostenberechnung für die ThinPrep™-Technologie (Kosten/Präparat)

	Kosten (€)
Verbrauchsmaterial (Bruttokosten) pro Präparat	11,31
Sonstige Kosten pro Präparat*	6,26
Porto pro Präparat	1,28
Gesamtkosten/Präparat	18,85

*Personalkosten, Raumkosten, Gerätekosten, andere Sachleistungen

Umlage der Anschaffungskosten eines Geräteprozessors

In den Berechnungen sind die Kosten für die zytologische Auswertung der Präparate nicht enthalten.

Für die Dünnschichtpräparationstechnologie AutoCytePrep™ wird eine Jahresuntersuchungskapazität von 50.000 Präparaten angegeben. Die Lebensdauer des Prozessors wird mit 7 Jahren und die Kosten für den Kauf eines Prozessors mit 83.229 € Brutto angegeben.

In Tabelle 39 wird die Kostenberechnung für die AutoCytePrep™-Technologie mit 3% Verzinsung und 3% Diskontierung für Kapitalkosten dargestellt. In den Berechnungen sind die Kosten für die zytologische Auswertung der Präparate nicht enthalten.

Tabelle 39: Kostenberechnung für die AutoCytePrep™-Technologie (Kosten/Präparat)

	Kosten (€)
Kapitalkosten pro Präparat	0,27
Instandhaltungskosten pro Präparat	0,05
Verbrauchsmaterial (Bruttokosten) pro Präparat	5,57
Sonstige Kosten pro Präparat*	0,84
Porto pro Präparat	1,28
Gesamtkosten/Präparat	8,01

*Personalkosten für Präparation, Raumkosten, sonstige Kosten.

In den Berechnungen sind die Kosten für die zytologische Auswertung der Präparate nicht enthalten.

Für das automatisierte Auswertungssystem AutoPap™ GS Screening wird eine Jahresuntersuchungskapazität von 50.000 Präparaten angegeben. Die Lebensdauer des Gerätes wird mit 5 Jahren eingegeben. Jährliche Leasing-Kosten für ein Gerät werden mit 99.143 € Brutto angegeben.

In Tabelle 40 wird die Kosten-Berechnung für die AutoPap™ GS Screeningtechnologie mit 3% Verzinsung und 3% Diskontierung für die Kapitalkosten dargestellt.

Tabelle 40: Kostenberechnung für die AutoPap™ GS Screeningtechnologie (Kosten/Präparat)

	Kosten (€)
Kapitalkosten pro Präparat	0,43
Instandhaltungskosten pro Präparat	0,54
Verbrauchsmaterial (Bruttokosten)	0,03
Sonstige Kosten pro Präparat*	2,00
Porto pro Präparat	1,28
Gesamtkosten/Präparat	4,28

*Personalkosten für Begutachtung/Nachbegutachtung, Raumkosten, sonstige Kosten.

II C 5. 4 Ökonomische Konsequenzen des Einsatzes neuer Technologien im Zervixkarzinom-Früherkennungsprogramm

II C 5. 4. 1 Internationale Kosten-Effektivitäts-Studien

Publikationen zu gesundheitsökonomischen Konsequenzen des Einsatzes von Dünnschichtpräparationen und automatisierten/halbautomatisierten Technologien zur Auswertung von Zervixzellabstrichen liegen bisher nur international aus den USA, Kanada, Australien und Großbritannien (AHTAC 1998; Brown & Garber 1999; Hutchinson et al. 2000; McCrory & Matchar 1999; Noorani et al. 1997; Payne et al. 2000; Raab et al. 1999b; Radensky & Mango 1998; Schechter 1996; Smith et al. 1999; Montz et al. 2001). Deutsche Publikationen wurden hierzu nicht gefunden (siehe Tabelle 41).

Tabelle 41: Internationale gesundheitsökonomische Studien

Autor	Jahr	Land	Perspektive	Auftraggeber	Technologie
(Schechter 1996)	1996	USA	Gesundheitsorganisation	Department of Community Medicine, Mt. Sinai School of Medicine und Leadership Consulting Associates, NY	PAPNET™
(Noorani et al. 1997).	1997	Ca	Kostenträger	Canadian Coordinating Office for Health Technology Assessment (CCOHTA)	AutoPap™ 300 QC, PAPNET™
(AHTAC 1998)	1998	Australien	Keine Angabe	AHTAC (Regierungsauftrag)	Automatisierte und halbautomatisierte Zervixscreening-Technologien
(Radensky & Mango 1998)	1998	USA	Kostenträger, modifiziert	McDermott, Will & Emery, FL und Neuromedical Systems Inc., Suffern, NY	Interactive Neural Network-Assisted Screening, PAPNET™
(Brown & Garber 1999)	1999	USA	Gesellschaft	Technology Evaluation Centre of the Blue Cross-Blue Shield Association health (insurance) plan	ThinPrep™, PAPNET™, AutoPap™ 300 QC
(McCrory & Matchar 1999)	1999	USA	Gesundheits-system	Agency for Healthcare Research and Quality (AHRQ)	Automatisierte und halbautomatisierte Zervixscreening-Technologien (100% Rescreening)
(Raab et al. 1999a)	1999	USA	Patient	Keine Angaben	Neue Technologien (automatisch und halbautomatisch)
(Smith et al. 1999)	1999	USA	Gesellschaft	Neopath Inc., Redmond, Washington	AutoPap™ Primary Screening System
(Hutchinson et al. 2000)	2000	USA	Kostenträger	Cytec Corporation, Boxborough, MA	ThinPrep™, AutoPap™ 300 QC, AutoPap™ Primary Screening System
(Payne et al. 2000)	2000	UK	Gesundheits-system	National Institute for Clinical Effectiveness (für NHS)	Liquid-based Screening
(Montz et al. 2001)	2001	USA	Keine Angabe	Finanziert durch Cytec Corporation, Boxborough, MA	Liquid-based Screening

Im folgenden werden die unten aufgeführten Studien nach Studienziel, Methoden, Ergebnissen, Autorenschlussfolgerung und Limitationen beschrieben.

Schechter, 1996 (Schechter 1996): Cost-Effectiveness of Rescreening Conventionally Prepared Cervical Smears by PAPNET™ Testing. (Zeitschriftenartikel)

Studienziel: Bewertung der Kosten-Effektivität des PAPNET™-Rescreeningverfahrens als Ergänzung zur konventionellen Zervixscreeningmethode.

Methoden: Das Markov-Modell von Muller et al. 1990 (Muller & Mandelblatt 1990) wurde adaptiert und verfolgte eine hypothetische Kohorte von 20- bis 64-jährigen Frauen, die nach dem US-amerikanischen Zensus von 1990 repräsentativ altersverteilt war und im Intervall von 2 Jahren gescreent wurde. Mortalitätsdaten wurden von „Vital Statistics of the United States“ bezogen. Die restlichen Basisparameter für das Modell wurden aus der Literatur (Sherlaw-Johnson et al. 1994; Eddy 1987; Eddy 1990) extrahiert. Für die Sensitivität des konventionellen Papanicolaou-Tests wurden 75% für LSIL und 85% für HSIL, sowie eine Spezifität von 98% festgelegt. Die inkrementelle Sensitivitätserhöhung für PAPNET™-Rescreening wurde auf 30% für LSIL und HSIL geschätzt (Herstellerangaben), die Spezifität auf 95%. Es wurde die Perspektive des Leistungsträgers eingenommen. Preise wurden der Krankenhaus-Gebührenverordnung nach ICD9-Klassifizierung, der Health Care Financing Administration und Fee for Service Plan entnommen. Es wurden direkte medizinische und nicht-medizinische Kosten für Diagnostik und Behandlung erhoben. Kosten für PAPNET™-Rescreening wurden vom Hersteller bezogen. Alle Kosten sind in 1994er

US\$ angegeben. Kosten und Gesundheitseffekte wurden mit 5% diskontiert. Es wurden folgende Strategien verglichen: 1. Konventioneller Papanicolaou-Test ohne Rescreening, 2. Konventioneller Papanicolaou-Test mit PAPNET™-Rescreening. Die Ergebnisse wurden in Kosten pro gewonnenem Lebensjahr berichtet.

Ergebnisse: Das inkrementelle Kosten-Effektivitäts-Verhältnis im Vergleich zu keinem Rescreening beträgt für ein Rescreening mit PAPNET™ 48.474 US\$ pro gewonnenem Lebensjahr bei einem Screeningintervall von 2 Jahren. Für ein 3-Jahres-Intervall würde sich eine IKEV von 25.185 US\$ pro gewonnenem Lebensjahr ergeben. Jährliches Screening mit PAPNET™-Rescreening war mit 113.078 US\$ nicht kosteneffektiv. In der Sensitivitätsanalyse verursachte die Variation der Behandlungskosten eine Änderung des inkrementellen Kosten-Effektivitäts-Verhältnisses von bis zu 16%. Eine Erhöhung der Sensitivität und Spezifität des Papanicolaou-Tests bzw. eine niedrigere Sensitivität von PAPNET™ erhöhte die IKEV drastisch.

Schlussfolgerungen der Autoren: Zusammenfassend stellen die Autoren fest, dass durch ein zusätzliches PAPNET™-Rescreening zum konventionellen Papanicolaou-Test die Mortalität und Morbidität des Zervixkarzinoms kostengünstig reduziert wer-

den kann. Ein Vergleich des Einsatzes von PAPNET™-Rescreening im 2-Jahres-Screening-Intervall mit anderen Präventionsprogrammen wie Mammografie, Östrogenbehandlung menopausaler Frauen und einmaliges Prostatakarzinom-Screening mit 50 Jahren zeigte eine deutlich höhere Kosten-Effektivität.

Limitierung der Studie: Die fehlende Darstellung absoluter Werte für Gesundheitseffekte und Kosten, sowie eine unzureichende Diskussion der Modellannahmen erschweren die Transparenz dieser Studie. Hauptlimitation für den Vergleich mit anderen Kosten-Effektivitäts-Studien zu Rescreening-Verfahren ist die gewählte Vergleichsstrategie des Screenings ohne Rescreening. Es ist anzunehmen, dass der größere Teil der Kosten-Effektivität dem konventionellen Pap-Test zuzuschreiben und das inkrementelle Kosten-Effektivitäts-Verhältnis im Vergleich zu einem Rescreening mit konventionellem Pap-Test sehr viel höher ist.

Noorani, 1997 (Noorani et al. 1997): Assessment of Techniques for Cervical Cancer Screening. (HTA-Bericht)

Studienziel: Beurteilung der Kosten-Effektivität der automatisierten Auswertungsmethoden AutoPap™ 300 QC und PAPNET™ als Rescreeningverfahren im Vergleich zum konventionellen Verfahren mit einem 10%igen manuellen Rescreening.

Methoden: Eine Entscheidungsbaumanalyse wurde eingesetzt, um die zu erwartenden direkten Gesamtlaborkosten pro zusätzlich gefundenen abnormalen Fall in einem Screeningzyklus zu ermitteln. Basisdaten für die Testcharakteristika wurden der Literatur entnommen. Die Analyse erfolgte aus der Perspektive eines Drittzahlers in Kanada. Es wurden direkte medizinische Kosten, die ausschließlich die Laborkosten berücksichtigten, verwendet. Klinische Kosten für Diagnostik, Therapie und Nachsorge wurden ausgeschlossen. Es wurden keine indirekten Kosten erhoben. Alle Kosten wurden aus der Literatur und durch einen Fragebogen an verschiedene Test-Hersteller ermittelt. Alle Kosten wurden in Dollar angegeben. Unklar bleibt, ob es sich um Can\$ oder US\$ handelt, da in Text und Tabellen beide Währungen aufgeführt werden. Ein Bezugsjahr wird nicht angegeben. Es wurde eine Anzahl von 4 Millionen Abstriche pro Jahr in der Analyse angenommen, welches der in Kanada geschätzten jährlichen Anzahl durchgeführter Untersuchungen entspricht. Kosten und Effekte wurden nicht diskontiert. Analyseergebnisse wurden in Kosten pro zusätzlich entdecktem abnormalen Fall angegeben. Der Zeithorizont umfasste einen Screeningzyklus von einem Jahr. Es wurden folgende Strategien verglichen: 1. Konventioneller Pap-Test ohne Rescreening, 2. Konventioneller Pap-Test mit 10%igem manuellen Rescreening, 3. Konventioneller Pap-Test mit AutoPap™ 300 QC Rescreening und 4. Konventioneller Pap-Test mit PAPNET™ Rescreening.

Ergebnisse: Im Vergleich zu keinem Rescreening ergaben sich folgende Kosten pro zusätzlichem abnormalen Fall: 251 \$ für ein 10%iges manuelles Rescreening, 418 \$ für AutoPap™-Rescreening und 810 \$ für PAPNET™-Rescreening. Im Vergleich zum manuellen Rescreening wurden inkrementelle Kosten pro zusätzlichem abnormalen Fall von 442 \$ für AutoPap™-Rescreening und 887 \$ für PAPNET™-Rescreening erzielt. PAPNET™-Rescreening im Vergleich zu AutoPap™-Rescreening erzielte eine IKEV von mehr als 10.000 Dollar pro zusätzlichem Fall. Das inkrementelle IKEV der beiden automatisierten Rescreeningmethoden im Vergleich zum manuellen Rescreening zeigte sich in Sensitivitätsanalysen sensitiv bei Variation der Kosten.

Schlussfolgerungen der Autoren: Die Autoren kommen zu dem Schluss, dass beide automatisierten Rescreeningverfahren effektiver als ein manuelles Rescreening jedoch auch kostenintensiver sind. Da bisher kein anerkanntes Standardmaß für einen Schwellenwert in Kosten-Effektivitäts-Analysen festgelegt wurde, hängt es von der Gesellschaft ab, wie viel sie bereit ist für einen zusätzlich entdeckten Fall zu bezahlen. Die Kosten pro zusätzlich gefundenen abnormalen Fall erscheinen für das AutoPap™-Rescreeningverfahren annehmbar.

Limitierung der Studie: Kardinale Schwäche dieser Studien liegt im Studiendesign. Es wurde kein Markov Model eingesetzt, welches die zugrundeliegende Erkrankung simuliert und die Berücksichtigung von Angaben zu Gesamtlebenszeitkosten und Überlebenszeit zulässt. Der Zeithorizont entsprach lediglich einem Screeningzyklus von einem Jahr. Kosten für Diagnostik, Therapie und Nachsorge wurden ausgeschlossen. Es wurden keine Gesundheitseffekte untersucht, sondern die zu erwartenden direkten Gesamtlaborkosten pro zusätzlich entdecktem Fall ermittelt. Die Definition eines abnormalen Falls blieb ebenso unklar wie die exakte Währung. In Text und Ergebnisdarstellung wurden sowohl US\$ als auch Can\$ angegeben. Kosten und Effekte wurden nicht diskontiert.

Ein Vergleich mit den anderen Kosten-Effektivitäts-Studien ist aus den aufgeführten Gründen nicht möglich.

AHTAC, 1998 (AHTAC 1998): Review of automated and semi-automated cervical screening devices. (HTA-Bericht)

Studienziel: Einschätzung der Kosten-Effektivität eines Einsatzes von automatisierten und halbautomatisierten Zervixscreeningverfahren im australischen Screening-Programm.

Methoden: Die Kosten-Effektivitäts-Analyse wurde mit einem modifizierten Flussdiagramm nach Braggett et al. 1993 durchgeführt, welches die potentielle Entwicklung geringgradiger und hochgradiger Dysplasien und die Wahrscheinlichkeit der Progression zum invasiven Karzinom berücksichtigte. Eine hypothetische Kohorte bestehend aus 5.022.752 australischen Frauen mit normalem Krebsrisiko im Alter von 20 und 74 Jahren (nach Australischem Zensus von 1996) durchlief einen Screeningzyklus von zwei Jahren. Aus einem systematischen Literaturreview und dem Victorian Cervical Cancer Registry wurden die Modellparameter entnommen. Es wurde eine 10% höhere LSIL- und 6% höhere HSIL-Entdeckungsrate angenommen. Die eingenommene Perspektive wurde nicht explizit dargestellt. Kosten für Diagnose und Behandlung von LSIL- und HSIL-Befunde wurden der Medicare Kostendatenbank und dem Australischen Casemix Report on Hospital Activity 1995-96 entnommen. Alle Kosten sind in AU\$ angegeben. Das Bezugsjahr wurde nicht genannt. Es wurde keine Diskontierung angegeben. Die Analyse ermittelte die Kosten pro verhindertem invasivem Karzinomfall nach einem einmaligen 2-Jahres Screeningintervall. Es wurden folgende Strategien untersucht: 1. konventionelles Screening und 2. zusätzlich Screening mit neuen Technologien.

Ergebnisse: Mit der Einführung einer automatischen oder halbautomatischen Screeningtechnologie in das australische Früherkennungsprogramm würden zusätzliche Kosten von 240.000 AU\$ pro potentiell entdecktem Krebsfall bei einem 2-Jahres Screeningintervall im Vergleich zum konventionellen Screeningverfahren entstehen.

Schlussfolgerungen der Autoren: Die Autoren schlussfolgern, dass die routinemäßige Einführung der neuen Technologien in das nationale Zervixkarzinom-Früherkennungsprogramm aus der Public-Health-Perspektive nicht kosteneffektiv ist. Es wurde die Durchführung einer australischen Studie empfohlen, um eine genauere Kostenschätzung sowohl für das konventionelle Screening als auch für die neuen Verfahren zu erhalten.

Limitierung der Studie: Hauptlimitation dieser Studie liegt im Modelldesign. Es wurde kein Markov-Modell, sondern ein für ein einmaliges Screening im 2-Jahres-Zyklus gerichtetes Flussdiagramm angewendet. Dieses simuliert nicht die zugrundeliegende Erkrankung und gibt weder Angaben zu Lebenszeitkosten noch Überlebenszeit. Es wurden keine direkten Gesundheitseffekte berücksichtigt, sondern die Kosten pro gefundenen abnormalen Befund in dem Screeningintervall bzw. pro potentiellen invasivem Karzinomfall berechnet. Da jedoch 1. ein abnormaler Befund nicht unmittelbar zwingend zu einem invasivem Karzinom führt und 2. bei einem nachfolgenden Screening entdeckt werden könnte, unterliegen die Analyseergebnisse einer hohen Unsicherheit. Der Einsatz der neuen Screeningtechnologien wird zusätzlich zum herkömmlichen australischen Vorsorgeprogramm bewertet und nicht als Ersatz. Die

Kostendatenerhebung ist nicht detailliert dargestellt. Die Kosten berücksichtigen keine Behandlungskosten des Zervixkarzinoms. Es wurden keine Diskontierungen der Kosten und Effekte angegeben, welches die neuen Technologien bevorteilt. Die Perspektive der ökonomischen Studie und die Kostendatenerhebung wurden nicht explizit beschrieben.

Ein Vergleich mit den anderen Kosten-Effektivitäts-Studien ist aus den aufgeführten Gründen nicht möglich.

Die Autoren diskutierten weiterhin die Ungewissheit hinsichtlich der Sensitivität der neuen Testverfahren, die sich aus der methodischen Qualität der Primärstudien ergibt, als Limitation. Es wurde das Fehlen eines adäquaten Referenzstandards und fehlende Angaben zu Laborbedingungen, wechselnde Schwellenwertangaben für ein positives Ergebnis und vorselektierte Studienpopulationen als Fehlerquellen in den Studien diskutiert.

Radensky, 1998 (Radensky & Mango 1998): Interactive Neural Network-Assisted Screening. An Economic Assessment. (Zeitschriftenartikel)

Studienziel: Gesundheitsökonomische Bewertung des Einsatzes von interactive neural network-assisted (INNA) Rescreeningverfahren in der Karzinomfrüherkennung.

Methoden: Die Kosten-Effektivitäts-Analyse basiert auf einem modifizierten neun Zustände umfassenden Markov-Modell von Eddy (Eddy 1990). Alle Annahmen des Originalmodells wurden beibehalten. Das Modell durchlief eine hypothetische Kohorte 20-jähriger Frauen, die im 3-Jahres-Intervall bis zum 75. Lebensjahr am Zervixscreening teilnahmen. Die Daten für die Sensitivität des INNA-Verfahrens wurden dem Review von Mango et al. 1998 (Mango & Radensky 1998) entnommen. Für die konventionelle Screeningmethode wurde eine Sensitivität von 85% angenommen. Die Kostenanalyse wurde aus der modifizierten Perspektive des Kostenträgers kombiniert mit Patientenzuzahlungen durchgeführt. Die Kosten für die gynäkologische Behandlung, Zytologie, Kolposkopie und Biopsie wurden aus den „Medicare fee schedules“ entnommen. Die INNA-Technologiekosten wurden vom Hersteller (PAPNET™) bereitgestellt. Die Behandlungskosten des Zervixkarzinoms stammen aus dem Originalmodell und wurden mit dem Consumer Price Index auf US\$1997 adaptiert. Kosten und Gesundheitseffekte wurden mit 3% diskontiert. Der Einsatz von INNA-Rescreeningtechnologien im Vergleich zum manuellen Rescreening negativer Befunde wurde evaluiert. Ergebnisse wurden in Kosten pro gewonnenem Lebensjahr dargestellt.

Ergebnisse: Es wurden drei verschiedene Referenzstandards für die Sensitivität des INNA-Tests, die mit Sensitivitäten im Bereich von 91% und 98% einhergingen, untersucht. Die inkrementellen Kosten-Effektivitäts-Verhältnisse im Vergleich zum manuellen Rescreening variierten entsprechend der Sensitivität zwischen 39.087 und 79.440 US\$ pro gewonnenem Lebensjahr. In der Sensitivitätsanalyse zeigte sich außerdem eine starke Abhängigkeit von der Sensitivität des konventionellen Papanicolaou-Tests. Bei einer Sensitivität von 90% stiegen die IKEV auf Werte zwischen 75.343 und 140.947 US\$ pro gewonnenem Lebensjahr an.

Schlussfolgerungen der Autoren: Die Autoren schlussfolgerten, dass sich bei einem Schwellenwert von 100.000 US\$ für die Kosten-Effektivität die Ausgaben für die INNA-Methode rechtfertigen lassen, um potentielle Zervixkarzinomfälle zu verhindern.

Limitierung der Studie: Die Basisdaten des Modells wurden dem Originalmodell von Eddy von 1990 entnommen. Kostendaten wurden nur aggregiert dargestellt. Preise und Mengengerüste wurden nicht explizit berichtet. Es wurden keine multivariaten Sensitivitätsanalysen durchgeführt. Es wurde nur ein Screeningintervall von drei Jahren untersucht.

Brown & Garber, 1999 (Brown & Garber 1999): Cost-effectiveness of 3 Methods to Enhance the Sensitivity of Papanicolaou Testing. (Zeitschriftenartikel)

Studienziel: Kosten-Effektivität der Technologien ThinPrep™, PAPNET™ und AutoPap™ im Vergleich zur konventionellen Papanicolaou-Methode.

Methoden: Die Kosten-Effektivitäts-Analyse basiert auf dem modifizierten neun Zustände umfassenden Markov-Modell von Eddy (Eddy 1990). Eine hypothetische Kohorte von 20- bis 65-jährigen für die US-Bevölkerung repräsentative Frauen durchläuft ein Screening mit Intervallen von ein, zwei, drei oder vier Jahren. Die Sensitivität des konventionellen Papanicolaou-Tests wird in Anlehnung an das „NIH's consensus panel“ (Eddy 1990) auf 80% geschätzt, die kumulative Sensitivität bei 10%igem Rescreening auf 81,6%. Die kumulative Sensitivität für ThinPrep™ wird mit 91,9% angegeben, die Gesamtsensitivität für ThinPrep™ mit einem 10%igen Rescreening auf 92,6% (Sheets et al. 1995). Die Sensitivität des AutoPap™ 300 QC Rescreening-Systems wird auf 77% bei 20% selektierter Rescreeningrate aller negativen Zellabstriche geschätzt. Bei Kombination mit initialem konventionellen Screening ergibt sich

eine kumulative Sensitivität von 95,4% (Colgan et al. 1997). Die kumulative Sensitivität für PAPNET™ wird mit 97% angegeben (Boon et al. 1994; Colgan et al. 1995; Jenny et al. 1997; Rosenthal et al. 1996). Da wenig bekannt ist über die Spezifität der neuen Tests, wurde eine Spezifität von 95,4% für alle Technologien angenommen. Die Studie wurde aus der gesamtgesellschaftlichen Perspektive durchgeführt. Es wurden direkte Kosten berücksichtigt. Testkosten wurden aus der Literatur, von Herstellern und Surveys in Pathologielaboren in North Carolina bezogen. Behandlungskosten wurden dem Originalmodell von Eddy 1990 entnommen und mittels Consumer Price Index auf 1996 adaptiert. Alle Kosten sind in 1996 US\$ angegeben. Kosten und Effekte wurden mit 3% diskontiert. Folgende Strategien wurden evaluiert: 1. Konventioneller Pap-Test mit 10%igem manuellen Rescreening, 2. ThinPrep mit 10%igem manuellen Rescreening, 3. Konventioneller Pap-Test mit AutoPap-Rescreening, 4. Konventioneller Pap-Test mit PAPNET-Rescreening. Die Ergebnisse wurden in Kosten pro gewonnenem Lebensjahr angegeben.

Ergebnisse: Im Vergleich zur konventionellen Papanicolaou- Methode mit 10%igem manuellen Rescreening führt der Einsatz von ThinPrep mit 10%igem manuellen Rescreening zu einer Kosten-Effektivität von 36.956 US\$ pro gewonnenem Lebensjahr bei einem 3-Jahres-Screeningintervall. ThinPrep plus 10%iges manuelles Rescreening wurde jedoch von der Strategie „Konventioneller Pap-Test mit AutoPap-Rescreening“ dominiert, welches eine inkrementelle IKEV von 16.259 US\$ pro gewonnenem Lebensjahr erzielte. Im Vergleich zum konventionellen Pap-Test mit 10%igem manuellen Rescreening bei gleichem Screeningintervall erzielte der Pap-Test mit PAPNET-Rescreening eine inkrementelle IKEV von 29.336 US\$ pro gewonnenem Lebensjahr. Die neuen Technologien waren bei kürzeren Screeningintervallen weniger kosteneffektiv. Eine Verbesserung der Kosten-Effektivität der neuen Technologien konnte in univariaten Sensitivitätsanalysen durch Erhöhung der Prävalenz der Erkrankung, eine niedrigere Sensitivität des konventionellen Papanicolaou-Tests und höhere Sensitivitäten der neuen Tests gezeigt werden.

Schlussfolgerungen der Autoren: Die Autoren schlussfolgern, dass ein Einsatz der neuen Methoden bei längeren Screeningintervallen sowie in Laboren mit hoher Falsch-Negativen-Rate kosteneffektiv sein kann, bei einem jährlichen Screening ist jedoch nur ein geringer Effekt zu erwarten. Kardinales Problem der Prävention des Zervixkarzinoms ist nach Aussagen der Autoren nicht die Sensitivität des Papanicolaou-Tests, sondern die Tatsache, dass einige Frauen nie gescreent werden.

Limitierung der Studie: Eingesetzte Sensitivitäten stammen von Studien mit erheblichen Unterschieden hinsichtlich Entnahmeanstrumente, Laborstandards und Erkrankungsprävalenzen. Die Diskussion der Modellannahmen ist unzureichend.

McCrorry, 1999 (McCrorry & Matchar 1999): Evaluation of Cervical Cytology. (HTA-Bericht)

Studienziel: Effektivität, Kosten und Kosten-Effektivität des Einsatzes von Technologien mit verbessertem Primärscreening (liquid-based Verfahren) sowie zusätzlich einem 100%igen automatisierten Rescreening negativer Befunde im Zervixscreeningprogramm der USA im Vergleich zur konventionellen Pap-Methode.

Methoden: Ein Markov-Modell mit 20 Gesundheitszuständen wurde entwickelt, in dem eine hypothetische Kohorte von 15 bis 85-jährigen für die US-Bevölkerung repräsentativen Frauen ein Screening im Intervall von ein, zwei oder drei Jahren durchläuft. Die Daten zu den Testcharakteristika wurden aus einer von den Autoren durchgeführten Metaanalyse in Verbindung eines systematischen Literaturreviews entnommen. Die Sensitivität des konventionellen Pap-Tests wurde mit 51%, die kumulativen Sensitivitäten der neuen Technologien mit 84% für die Primärscreening-Technik und 80% für die Rescreening-Technik angegeben. Es wurde für alle Technologien dieselbe Spezifität von 97% eingesetzt. Da die Evidenz der Testcharakteristika unbestimmt war, wurde eine Schwellenwertanalyse durchgeführt, um den Schwellenwert für die Sensitivität und Spezifität zu ermitteln, bei dem das Kosten-Effektivitäts-Verhältnis unterhalb 50.000 US\$ pro gewonnenem Lebensjahr fällt. Die Analyse erfolgte aus der Perspektive des US-Gesundheitssystems. Es wurden direkte medizinische Kosten evaluiert, die sich auf Screening, Diagnostik und Therapie des Zervixkarzinoms und seiner Vorstadien beziehen. Indirekte medizinische Kosten wurden nicht erhoben. Die Kostendaten wurden unterschiedlichen US-amerikanischen Gesundheitsdiensten und Versicherungen entnommen. Außerdem wurden Durchschnittszahlungen nach diagnosis related groups (DRG's) angegeben. Alle Kosten wurden in 1997 US\$ angegeben. Kosten und Effekte wurden mit 3% diskontiert. Alle Modellannahmen wurden explizit und ausführlich dargelegt. Folgende Strategien wurden evaluiert: 1. "Kein Screening", 2. Konventionelles Pap-Screening mit 10%igem manuellen Rescreening, 3. Verbessertes Primärscreening mit einem 10%igem manuellen Rescreening, 4. konventionelles Pap-Screening mit einem 100%igen automatisierten Rescreening. Die Ergebnisse wurden in Kosten pro gewonnenem Lebensjahr angegeben.

Ergebnisse: Unter der Annahme, dass die Falsch-Negativen-Ratio (FNR) durch den Einsatz der verbesserten Primärscreening-Technik um 60% reduziert werden kann, wurde mit dieser Strategie im Vergleich zum konventionellen Verfahren bei einem 3-Jahres-Intervall eine inkrementelle IKEV von 22.010 US\$ pro gewonnenem Lebensjahr und ein Gewinn von 2,2 Lebenstagen erzielt. Das 100%ige automatisierte Rescreening nach einem konventionellen Screening ergab eine inkrementelle IKEV von 30.507 US\$ pro gewonnenem Lebensjahr und einem Gewinn von 2,01 Lebenstagen.

Längere Intervalle ergaben günstigere IKEV, während kürzere Intervalle von zwei Jahren und weniger nicht kosteneffektiv waren. Die Ergebnisse waren nicht sensitiv gegenüber Annahmen zur Zervixkarzinominzidenz, zu Technologiekosten, diagnostischen Strategien bei auffälligen Screeningergebnissen, Alter bei erstem Screening oder anderen getesteten Variablen, jedoch hinsichtlich der Sensitivität der Technologien.

Schlussfolgerungen der Autoren: Die Autoren schlussfolgern, dass der Einsatz der neuen Technologien unter optimalen Bedingungen im 3-Jahres-Screeningintervall im Vergleich zur konventionellen Pap-Methode eine akzeptable Kosten-Effektivität erreichen kann. Jedoch sollte beachtet werden, dass der Nutzen gemessen an gewonnener Lebenszeit und der Reduktion der Krebsinzidenz gering ist, da die meisten der entdeckten zytologischen Abnormalitäten geringgradige Dysplasien darstellen. Außerdem wird eine große Ungewissheit hinsichtlich Sensitivität und Spezifität der neuen Verfahren aufgezeigt. Studien zur reliablen Einschätzung der Testcharakteristika und des Einflusses der neuen Technologien auf die Lebensqualität werden empfohlen.

Limitierung der Studie: Im Unterschied zu den anderen Studien wird ein verhältnismäßig niedriger Wert für die Sensitivität des konventionellen Papanicolaou-Tests angegeben.

Raab, 1999 (Raab et al. 1999b): The Cost-Effectiveness of the Cytology Laboratory and New Cytology Technologies in Cervical Cancer Prevention. (Zeitschriftenartikel)

Studienziel: Untersuchung der Änderung der Kosten-Effektivität des Zervixscreenings durch die Einführung neuer Technologien.

Methoden: Die Kosten-Effektivitäts-Analyse wurde mit Hilfe eines entscheidungsanalytischen Modells durchgeführt, welches Erkrankungsrisiken und Kosten der Behandlung von diagnostizierten abnormalen Läsionen und Karzinomen berücksichtigt. Es wurde ein Referenzfall für eine 30-jährige, weiße Frau erstellt. Daten zur Krankheitsentwicklung stammten aus der Literatur, sowie der SEER-Datenbank und dem National Center for Health Statistics. Kostendaten wurden von den University of Iowa Hospitals and Clinics sowie der medizinischen Literatur entnommen. Die Analyse erfolgte aus der Perspektive des Patienten. Es wurden direkte medizinische und nicht medizinische Kosten, sowie Screening- und Interpretationszeiten von Zytologieassistenten und –pathologen berücksichtigt. Alle Kosten wurden in 1997 US\$ berichtet. Die Lebenserwartung wurde mit 5% diskontiert, während die Kosten nicht diskontiert

wurden. Es wurden folgende Strategien evaluiert: 1. Kein Screening, 2. Screening mit der konventionellen Pap-Methode. Der Einfluss der neuen Technologien wurde in einer Sensitivitätsanalyse untersucht, bei der ermittelt wurde, wie viel zusätzliche HSIL-Befunde bei welchen zusätzlichen Testkosten der neuen Technologien erforderlich sind, um kosteneffektiv zu sein. Die Ergebnisse wurden in Kosten pro gewonnenem Lebensjahr angegeben.

Ergebnisse: Verglichen mit der Strategie kein Screening, ergab sich für den konventionellen Zellabstrich eine inkrementelle IKEV von 19.655 US\$ pro diskontiertem gewonnenem Lebensjahr. In der Sensitivitätsanalyse zeigte sich, dass die neuen Technologien bei Mehrkosten von 10 US\$ kosteneffektiv, d.h. unterhalb 50.000 US\$ pro gewonnenem Lebensjahr waren, wenn mindestens 236 hochgradige Dysplasien pro 10.000 Frauen zusätzlich entdeckt werden. Ist die Gesellschaft bereit, mehr für ein gewonnenes Lebensjahr auszugeben, sind die neuen Technologien auch bei einer geringeren HSIL-Detektionsrate kosteneffektiv.

Schlussfolgerungen der Autoren: Die Autoren schlussfolgern, dass bei einer angenommenen Kosten-Effektivitäts-Grenze von 50.000 US\$ die herkömmliche Zervixscreening-Methode selbst unter Berücksichtigung der Behandlungskosten kosteneffektiv ist. Der Einsatz der neuen Technologien ist nach Ansicht der Autoren nur dann kosteneffektiv, wenn diese eine entsprechend hohe Rate mit hochgradigen Dysplasien entdecken. Jedoch wurden in der Literatur (Patten S.F. et al. 1996b; Lee J.S. et al. 1998; Mango 1996b; O'Leary et al. 1998) derartige hohe HSIL-Detektionsraten für die neuen Technologien bisher nicht beschrieben. Die Autoren empfehlen zusätzliche prospektive Studien mit neuen Zervixscreeningverfahren bei unterschiedlichen Populationen.

Limitierung der Studie: In der Basisanalyse wurde Screening versus keinem Screening mit dem Pap-Test evaluiert. Der Einfluss der neuen Technologien wurde nur in der Sensitivitätsanalyse untersucht, welches einen Vergleich mit den anderen Studien erschwert. Eine Diskontierung erfolgte nur für die Effekte, nicht für die Kosten. Die Kostendatenerhebung wurde nicht explizit dargestellt.

Smith, 1999 (Smith et al. 1999): Economic Impact of Automated Primary Screening for Cervical Cancer. (Zeitschriftenartikel)

Studienziel: Ökonomische Bewertung des Einsatzes des AutoPap™ Primär-Screeningsystems in der klinischen Praxis im Vergleich zum herkömmlichen Pap-Test.

Methoden: Die Kosten-Effektivitäts-Analyse wurde mit Hilfe eines sieben Zustände umfassenden Markov-Modells in Anlehnung an Eddy (Eddy 1990) durchgeführt. Eine hypothetische Kohorte bestehend aus 18-jährigen für die US-Bevölkerung repräsentativen Frauen, die im 3-Jahres-Rhythmus gescreent wurden, durchlief das Modell. Angaben zu Testcharakteristika wurden aus der Literatur entnommen. Die kumulative Sensitivität des Papanicolaou-Tests wurde mit 81,6% für LSIL-Schwellenwert (Brown & Garber 1999), die des AutoPap™-Primär-Screeningsystems mit 91% (Wilbur D.C. et al. 1998), die kumulative Spezifität des Papanicolaou-Tests mit 94% (Wilbur D.C. et al. 1998) für ASCUS-Schwellenwert und die des AutoPap™-Tests mit 95% (Wilbur D.C. et al. 1998) angegeben. Die Mortalitätsdaten wurden der SEER-Datenbank bzw. dem „National Center for Health Statistics“ entnommen. Es wurde die gesamtgesellschaftliche Perspektive eingenommen. Es wurden direkte medizinische Kosten erhoben. Die Kostendaten zu Diagnostik und Behandlung des Zervixkarzinoms stammten aus Medicare-Veröffentlichungen. Die Kosten für die AutoPap™-Technologie wurde vom Hersteller bezogen. Alle Kosten wurden in 1997er US\$ berichtet. Kosten und Gesundheitseffekte wurden mit 3% diskontiert. Folgende Strategien wurden verglichen: 1. Konventionelles Screening mit 15% manuellem Rescreening und 2. AutoPap™-Screening mit 15%igem AutoPap™-Rescreening. Die Ergebnisse wurden in Kosten pro gewonnenem Lebensjahr berichtet.

Ergebnisse: Das Ergebnis der Analyse zeigte für das AutoPap™ Primary Screening System einen Gewinn von zusätzlichen 13,1 Lebenstagen pro Frau bei einem 3-Jahres-Screeningintervall und im Vergleich zur konventionellen Methode eine Kostenreduktion von 975 US\$ pro gewonnenem Lebensjahr. In den Sensitivitätsanalysen zeigte sich, dass das Basismodell sensitiv auf Veränderungen der Kosten des AutoPap™-Tests und der Zervixkarzinom-Prävalenz reagierte.

Schlussfolgerungen der Autoren: In der Zusammenfassung stellen die Autoren fest, dass automatisierte Zervixscreening-Techniken signifikant gesundheitliche Ergebnisse verbessern und Kosten reduzieren können. Sie geben aber zu bedenken, dass die aktuellen Kosten für die neuen Methoden möglicherweise höher sein könnten, was bei einer flächendeckenden Einführung in Präventionsprogramme beachtet werden sollte.

Limitierung der Studie: Die Ergebnisse dieser Studie sind widersprüchlich zu allen anderen Kosten-Effektivitäts-Studien, in denen die neuen Technologien immer mit zusätzlichen Kosten pro gewonnenem Lebensjahr verbunden sind. Es wurde kein systematischer Review zur Extraktion der Daten für die Testcharakteristika durchgeführt. Werte für die Sensitivität und Spezifität stammten aus unterschiedlichen Studien, die unterschiedliche Testschwellenwerte benutzten (Sensitivität LSIL, Spezifität ASCUS). Die Studie, welche die Daten zur Sensitivität von AutoPap™ lieferte (Wilbur

D.C. et al. 1998), analysierte eine Hochrisiko-Gruppe (das Haupteinsatzgebiet dieser neuen Technologie), während Smith et al. eine „normale“ Studienpopulation untersuchten. Es bleibt unklar, warum die Sensitivität des Papanicolaou-Tests aus einer anderen Studie entnommen wurde. Alle Testcharakteristika hätten aus derselben Studie extrahiert werden können. Multivariate Sensitivitätsanalysen wurden nicht durchgeführt.

Hutchinson, 2000 (Hutchinson et al. 2000): Clinical and Cost Implications of New Technologies for Cervical Cancer Screening: The Impact of Test Sensitivity. (Zeitschriftenartikel)

Studienziel: Vergleich der verfügbaren Techniken zum Zervixkarzinom-Screening unter Berücksichtigung verschiedener neuer Techniken (AutoPap™ 300 QC, AutoPap™ Primary Screening System, ThinPrep™), im Rahmen des aktuellen Screeningprogramms in den USA.

Methoden: Die Kosten-Effektivitäts-Analyse wird mit Hilfe eines Markov-Modells durchgeführt. Die hypothetische Kohorte besteht aus 100.000 Frauen im Alter von 20 Jahren, die bis zu einem Alter von 65 Jahren gescreent werden. Es wurden Screeningintervalle von 1, 2, 3, 5, und 10 Jahren eingesetzt. Epidemiologische Modelldaten entstammen dem National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) bzw. Centers for Disease Control and Prevention. Modellparameter zur Erkrankungsprogression wurden der Literatur entnommen. Alle Modellannahmen wurden detailliert beschrieben. Es wurde die Kostenträger-Perspektive eingenommen. Es wurden direkte medizinische Kosten betrachtet. Kostendaten wurden auf Gebührenebene oder Fee for Service der Medicare Datenbank und der Guideline for Fee Schedule and Management entnommen. Die Testkosten wurden vom Hersteller bezogen. Alle Kosten wurden in 1997er US\$ angegeben. Kosten und Effekte wurden mit 3% diskontiert. Es wurden folgende Strategien verglichen: 1. Konventionelles Pap-Screening mit 10%igem manuellen Rescreening und 2. Konventionelles Pap-Screening mit AutoPap QC 300 Rescreening, 3. AutoPap Primärscreening mit 10%igem manuellen Rescreening, 4. ThinPrep Screening mit 10%igem manuellen Rescreening. Die Ergebnisse wurden in Kosten pro gewonnenem Lebensjahr berichtet.

Ergebnisse: Im Vergleich zur konventionellen Screeningmethode erzielte ThinPrep™ (jeweils mit 10%igem manuellen Rescreening) inkrementelle Kosten-Effektivitäts-Verhältnisse von 3.045 bis 120.129 US\$ pro gewonnenem Lebensjahr bei Screeningintervallen von 10 Jahren bis zu jährlichem Screening. Kosteneffektiv bei einem Schwellenwert von 50.000 US\$ pro gewonnenem Lebensjahr ist das Screening mit

ThinPrep™ bei Screeningintervallen von 2 oder mehr Jahren. AutoPap™ 300 QC und AutoPap™ Primary Screening System weisen in allen Intervallen doppelte bis dreifache IKEV auf, sind jedoch ab einem Screeningintervall von drei oder mehr Jahren auch kosteneffektiv. In der Sensitivitätsanalyse wurden Kosten basierend auf dem National Hospital Interview Survey eingesetzt, die zu sehr viel höheren inkrementellen IKEV der neuen Technologien im Vergleich zur konventionellen Screeningmethode führten. In einer Population, die regulär mindestens alle drei Jahre gescreent wurde, war keine der Technologien kosteneffektiv.

Schlussfolgerungen der Autoren: Die Autoren schlussfolgern, dass durch die erhöhte Sensitivität der neuen Technologien eine kosteneffektive Reduktion neuer Zervixkarzinomfälle erreicht werden kann und damit eine realistische Lösung bei der Bekämpfung des Zervixkarzinoms gefunden wurde.

Limitierung der Studie: Effekte wurden in der Ergebnispräsentation nicht absolut pro Kopf dargestellt. Univariate Sensitivitätsanalysen wurden nicht für die Testcharakteristika beschrieben. Multivariate Sensitivitätsanalysen wurden nicht durchgeführt. Die Diskussion der Modellannahmen ist unzureichend.

Payne, 2000 (Payne et al. 2000): Liquid-based cytology in cervical screening: a rapid and systematic review. (HTA-Bericht)

Studienziel: Einfluss des Einsatzes von liquid-based Technologien im Vergleich zur konventionellen Pap-Methode auf Zervixkrebsinzidenz, Mortalität, Kosten und Kosten-Effektivität.

Methoden: Die Kosten-Effektivitäts-Analyse wurde mit Hilfe des Markov-Modells nach Sherlaw-Johnson et al. 1994 (Sherlaw-Johnson et al. 1994) durchgeführt. Eine hypothetische Kohorte von 100.000 Frauen wurde vom 18. bis 95. Lebensjahr verfolgt. Frauen im Alter von 21 bis 64 Jahren wurden einem Zervixkrebsscreening im Intervall von zwei, drei oder fünf Jahren unterzogen. Zykluslänge des Modells betrug 6 Monate. Alle Parameter der Erkrankungsprogression und Basisparameter der Testcharakteristika wurden dem Originalmodell entnommen. Mortalitätsdaten wurden aus Sterbetafeln von 1992-1994 des Governments Actuary's Department England/Wales bezogen. Es wurde eine 85%ige Teilnehmerate am Screening angenommen, 15% der Kohorte wurden nie gescreent. Die verwendeten Werte für die Testsensitivität wurden abhängig vom Grad der histologischen Veränderung angenommen. Die Wahrscheinlichkeit eines Papanicolaou-Testergebnisses mild/grenzwertig wird mit 39% für CIN 1, 22% für CIN 2, 50% für CIN 3 und 60% für invasives Karzinom und die eines Papanicolaou-Testergebnisses von mode-

rat/schwerwiegend mit 4% für CIN 1, 15% für CIN 2, 50% für CIN 3 und 40% für invasives Karzinom angegeben. Für liquid-based Technologien wurde eine marginale Sensitivitätserhöhung von 15% für CIN 1 und CIN 2, sowie 2% für CIN 3 und höher basierend auf einem systematischen Literaturreview angegeben. Die Spezifität wurde für beide Technologien auf 98% geschätzt. Der Anteil inadäquater Abstriche wurde mit 3% für liquid-based und 9% für den konventionellen Pap-Test basierend auf Daten aus Großbritannien angegeben. Die Analyse nahm die Perspektive des Gesundheitssystems ein. Es wurden ausschließlich direkte medizinische Kosten für Screening, Diagnostik und Behandlung berücksichtigt. Langzeitbehandlungen, adjuvante und palliativ-medizinische Maßnahmen wurden aus der Kostenberechnung ausgeschlossen. Kosten wurden aus der NHS CSP-Analyse von 1994 bezogen und auf 1999iger Werte adaptiert. Die Diskontrate wurde für die Kosten mit 6% und für die medizinischen Effekte mit 1,5% angegeben. Es wurden die Modellstrategien 1. konventionelle Methode versus 2. liquid-based Technologie verglichen. Die Ergebnisse wurden in Kosten pro verhindertes invasives Karzinom und in Kosten pro gewonnenem Lebensjahr angegeben.

Ergebnisse: Bei einem 3-Jahres-Screeningintervall führt der Einsatz von liquid-based Technologien im Vergleich zur konventionellen Methode zu inkrementellen Kosten von 2.723 £ pro vermiedenem invasiven Karzinomfall und inkrementellen Kosten von 2.522 £ pro gewonnenem Lebensjahr. Geringere IKEV werden bei längeren Screeningintervallen gezeigt. In der Sensitivitätsanalyse wurde eine Diskontierung von 3% für Kosten und Effekte durchgeführt. Diese Ergebnisse sind mit anderen Studien besser vergleichbar. Es wurde eine inkrementelle IKEV von 25.454 £ pro gewonnenem Lebensjahr für ein Screeningintervall von 3 Jahren angegeben. Sensitivitätsanalysen zeigten, dass die Ergebnisse sensitiv in Bezug auf Variationen in der Diskontierung und den inkrementellen Testkosten sowie in den marginalen Test-Spezifitäten und der Rate inadäquater Abstriche sind.

Schlussfolgerungen der Autoren: Die Autoren resümieren, dass die Einführung der liquid-based Technologie im 3-Jahres-Screeningintervall im akzeptablen Bereich der Kosten-Effektivität liegen könnte. Allerdings warnen sie gleichermaßen vor einer Überbewertung der Ergebnisse, da in der Analyse einige Parameter einer Ungewissheit unterliegen und in den Gesundheitseffekten der beiden Strategien nur kleine Unterschiede bestehen. Die Autoren empfehlen weitere Studien durchzuführen, insbesondere unter kritischer Betrachtung der verwendeten Modelldaten und –annahmen.

Limitierung der Studie: Hauptlimitation dieses HTA-Berichtes ist die unterschiedliche Diskontierung von Kosten und Effekten in der Basisfall-Analyse. Bei Diskontierung von Kosten und Effekten mit gleicher Diskontrate sind sehr viel höhere Kosten pro

gewonnenem Lebensjahr zu erwarten. Es ist jedoch anzumerken, dass dieses in der Sensitivitätsanalyse und im Appendix des Berichts ausreichend dargestellt wird. Als weitere Limitation ist anzusehen, dass Inzidenz von präinvasiven Läsionen und die Progressionsraten nicht altersspezifisch ins Modell eingegangen sind, welches zu einer Unterschätzung des Screeningeffektes generell führt. Hervorzuheben ist, dass im Unterschied zu anderen Studien die Teilnahmerate am Screening mit 85% in die Analyse einging.

Montz et al., 2001 (Montz et al. 2001): Impact of Increasing Papanicolaou Test Sensitivity and Compliance: Modeled Cost and Outcomes Analysis. (Zeitschriften-Artikel)

Studienziel: Einfluss der Erhöhung der Teilnahmerate am Screening und des Einsatzes von liquid-based Zytologie in Populationen mit bekanntem Teilnahmeverhalten auf die Zervixkrebsinzidenz und Kosten-Effektivität.

Methoden: Die Kosten-Effektivitäts-Analyse wurde mit Hilfe des Markov-Modells nach Hutchinson et al. 2000 (Hutchinson et al. 2000) durchgeführt. Eine hypothetische Kohorte von 100.000 Frauen wurde vom 20. bis 80. Lebensjahr verfolgt. Die Regressionsrate aller CIN-Läsionen wurde auf 65% über 6 Jahre bei Frauen im Alter von 20-34 Jahre und auf 40% über 6 Jahre bei Frauen ab einem Alter von 35 Jahren angegeben. Die Regressionsrate von CIS wurde mit 35% über 6 Jahre angegeben. Die Progression wurde über zwei Stadien von CIN über CIS nach invasiven Karzinom modelliert, wobei die Progressionsraten basierend auf Inzidenzraten des jeweiligen Zustands modelliert wurden. Mortalitätsdaten wurden vom National Center for Health Statistics bezogen. Vom National Cancer Institute's SEER Cancer Statistics Review wurden die Daten zur Zervixkarzinominzidenz, Mortalitätsraten, sowie 5-Jahres-Überlebensraten bei Zervixkarzinom bezogen. Altersspezifische Hysterektomieraten wurden vom Center for Disease Control bezogen. Als Basisfall wurde eine falsch-negative Befundrate von 49% für das Papanicolaou-Verfahren und von 27% für das liquid-based Verfahren angegeben. Es wurden ausschließlich direkte medizinische Kosten für Screening, Diagnostik und Behandlung berücksichtigt. Es wurden die Leistungen berichtet von der American Medical Association zugrunde gelegt und die Kosten entsprechend der Gebührenverordnung für Ärzte berechnet. Alle Kosten sind in 1997iger US\$ angegeben. Die Diskontrate wurde mit 3% für medizinische Effekte und Kosten angegeben. Es wurden die Modellstrategien 1. konventioneller Papanicolaou-Test mit 10% Rescreening negativer Befunde versus 2. liquid-based Technologie miteinander verglichen und zwar jeweils bei A. einer von Frauen selbstberichteten Teilnahmerate, B. einer Teilnahmerate angestrebt nach „Healthy People 2000“ und C. einer Teilnahmerate angestrebt nach „Healthy People 2010“. Die Teilnahmeraten wurden unterteilt in a. niemals teilgenommen, b. in den letzten 3 Jahren nicht teilge-

nommen, c. in den letzten 3 Jahren mindestens einmal teilgenommen. Die selbstberichteten Teilnahmeraten entstammen dem Bericht des Behavioral Risk Factor Surveillance Systems von 1997. Die Teilnahmerate angestrebt nach „Healthy People 2000“ wurde auf a. 5%, b. 10% und c. 85%, die Raten angestrebt nach „Healthy People 2010“ auf a. 3%, b. 7% und c. 90% angegeben. Es wurde die Annahme gemacht, dass die in Gruppe b. 50% der Frauen alle 5 Jahre und 50% alle 10 Jahre und in Gruppe c. jeweils ein Drittel der Frauen jährlich, alle 2 Jahre bzw. alle 3 Jahre gescreent wurden.

Die Ergebnisse wurden jeweils in Kosten pro gewonnenem Lebensjahr angegeben.

Ergebnisse: Im Vergleich zum Papanicolaou-Test führt der Einsatz von liquid-based Zytologie zu inkrementellen Kosten von 15.296 US\$ pro gewonnenem Lebensjahr bei selbstberichteter Teilnahmerate, 16.796 US\$/LJ bei einer Teilnahmerate angestrebt nach „Healthy People 2000“ und 20.424 US\$/LJ bei einer Teilnahmerate angestrebt nach „Healthy People 2010“. Bei einer Variation der Falsch-Negativ-Rate der liquid-based Zytologie auf 40% blieben die inkrementellen Kosten-Effektivitäts-Verhältnisse unter 50.000 US\$/LJ. Bei einer Erhöhung der Teilnahmeraten nach den „Healthy People“ Agenden erzielte der Einsatz der liquid-based Technologie günstigere inkrementelle Kosten-Effektivitäts-Verhältnisse als die Papanicolaou-Methode, wenn der Papanicolaou-Test bei selbstberichteter Teilnahmerate als Referenz diente.

Schlussfolgerungen der Autoren: Die Autoren resümieren, dass die Einführung der liquid-based Technologie im akzeptablen Bereich der Kosten-Effektivität liegen könnte. Insbesondere in Hoch-Risikogruppen kann die liquid-based Technologie die kosteneffektivste Screeningtechnologie sein. Der Einsatz von liquid-based Technologie in Verbindung mit einer Motivation zur Erhöhung der Teilnahmerate der Frauen kann als kosteneffektiver angesehen werden als eine Teilnahmeratenerhöhung allein.

Limitierung der Studie: Diese Studie untersuchte als bisher einzige den Einsatz der liquid-based Technologie bei unterschiedlichem Teilnahmeverhalten. Leider wurde jedoch die selbstberichtete Teilnahmerate, die auch als Referenzfall für das derzeitige Teilnahmeverhalten herangezogen wurde, nicht ausreichend detailliert beschrieben. Es ist nicht ersichtlich um wie viel Prozent die Teilnahmerate erhöht wurde und in welchen Zeitintervallen die Frauen derzeit am Screening teilnehmen. In dieser Studie wurde das Screeningintervall variabel durch das Screeningverhalten in der jeweiligen Strategie bestimmt. Es wurden absolute weder für die Referenzstrategie noch für alle anderen Strategien Kosten und Effekte berichtet. Es wurde keine Anga-

be zur Spezifität der Tests gemacht. Das Model unterscheidet nicht zwischen verschiedenen Zervixkarzinomstadien.

Zusammenfassung

Alle eingeschlossenen ökonomischen Studien weisen Limitierungen auf. Keine der Studien hat die Gesundheitseffekte mit Präferenzen bewertet und bei der Kostenerhebung indirekte Kosten berücksichtigt (Tabelle 42). Kosten für die neuen Technologien berücksichtigten in der Regel keine Personalschulungskosten, Lagerungskosten und Transportkosten. Die Werte für Sensitivität und Spezifität (Tabelle 45 und Tabelle 46) der Testverfahren unterlagen einer großer Variabilität und wurden oft infolge fehlender Evidenz geschätzt. Zusätzlich wurde häufig angenommen, dass sich die marginale Erhöhung der Sensitivität der neuen Technologie bei allen Schwellenwerten (CIN 1-III, LSIL, HSIL) gleich ausprägt.

II C 5. 4. 2 Vergleich der Kosten-Effektivität der neuen Technologien

Es folgt ein Vergleich der Ergebnisse der internationalen Kosten-Effektivitäts-Studien. Auf Grund der im vorherigen Kapitel dargelegten Diskrepanzen in den Studiendesigns und mangelnder Ergebnispräsentation wurden hiervon die folgenden sechs Studien ausgeschlossen:

AHTAC (AHTAC 1998), Noorani (Noorani et al. 1997), Smith (Smith et al. 2000), Raab (Raab et al. 1999b), Schechter (Schechter 1996), Montz (Montz et al. 2001)

Primärscreening-Technologien

Tabelle 43 gibt eine Übersicht über die inkrementellen Kosten, Gesundheitseffekte und Kosten-Effektivität von Primärscreening-Technologien im Vergleich zur konventionellen Papanicolaou-Technologie in den evaluierten Studien. Es wurden Ergebnisse gegenübergestellt mit vergleichbarer Diskontierung der Kosten und Effekte. Inkrementelle Kosten wurden in US\$ angegeben und wo notwendig mit dem Währungskonversionsfaktor für den Gesundheitssektor (OECD Health Data 2000) konvertiert. Alle inkrementellen Kosten und Effekte wurden im Vergleich zur konventionellen Papanicolaou-Methode dargestellt.

Das inkrementelle Kosten-Effektivitäts-Verhältnis der Primärscreening-Technologien im Vergleich zur konventionellen Papanicolaou-Technologie liegt im Bereich von 10.400 - 37.000 US\$/LJ bei einem Screeningintervall von 3 Jahren und kann als kosteneffektiv eingeschätzt werden, wenn der in der Literatur weithin akzeptierte Schwellenwert von 50.000 US\$/LJ als Maß herangezogen wird. Bei kürzeren Screeningintervallen werden überwiegend inkrementelle Kosten-Effektivitäts-Verhältnisse über 50.000 US\$/LJ berichtet.

Rescreening-Technologien

Tabelle 44 gibt eine Übersicht über die inkrementellen Kosten, Gesundheitseffekte und Kosten-Effektivität von Rescreening-Technologien im Vergleich zur konventionellen Papanicolaou-Technologie in den evaluierten Studien. Es wurden Ergebnisse gegenübergestellt mit vergleichbarer Diskontierung der Kosten und Effekte. Inkrementelle Kosten wurden in US\$ angegeben und wo notwendig mit dem Währungskonversionsfaktor für den Gesundheitssektor (OECD Health Data 2000) konvertiert.

Alle inkrementellen Kosten und Effekte wurden im Vergleich zur konventionellen Papanicolaou-Methode dargestellt.

Das inkrementelle Kosten-Effektivitäts-Verhältnis der Rescreening-Technologien im Vergleich zur konventionellen Technologie liegt im Bereich von 16.100 - 61.800 US\$/LJ bei einem Screeningintervall von 3 Jahren und kann als kosteneffektiv eingeschätzt werden. Bei kürzeren Screeningintervallen werden überwiegend inkrementelle Kosten-Effektivitäts-Verhältnisse über 50.000 US\$/LJ berichtet.

Tabelle 43: Inkrementelle Kosten und Effekte von neuen Primärscreeningverfahren im Vergleich zum konventionellen Pap-Test

Quelle	Land	Währung / Jahr	Technologie	Diskontrate	Screening- intervall (Jahre)	Inkrementelle Kosten (US\$)	Inkrementelle Effekte (LT)	Inkrementelle KE-Ratio (US\$/LJ)
McCroy & Matchar1999	USA	US\$ 1997	Primärscreening Technologie	3%	1	298	0,6	173.484
					2	178	1,4	45.265
					3	132	2,2	22.010
Payne et al. 2000	UK	(UK£ 1999) US\$ 1996 ^c	Liquid-based Technologie	3%*	2	K. A.	K. A.	52.576 (£128.234)*
					3	K. A.	K. A.	10.436 (£25.454)*
Brown & Garber 1999	USA	US\$ 1996	ThinPrep™	3%	1	239	0,24	363.479
					2	120	0,47	93.192
					3	81	0,80	36.956
Hutchinson et al. 2000	USA	US\$ 1997	ThinPrep™	3%	1	129	0,393	119.809
					2	54	0,687	28.690
					3	30	0,908	12.059
Hutchinson et al. 2000	USA	US\$ 1997	AutoPap™	3%	1	50	0,085	214.706
					2	23	0,144	58.299
					3	15	0,188	29.122

LT: gewonnene Lebensstage, LJ: gewonnene Lebensjahre, K.A.: Keine Angabe.

* Sensitivitätsanalyse; ^c Konversion mit dem Konversionsfaktor für den Gesundheitssektor (OECD Health Data 2000).

Tabelle 44: Inkrementelle Kosten und Effekte von neuen Rescreeningverfahren im Vergleich zum konventionellen Pap-Test

Quelle	Land	Währung / Jahr	Technologie	Diskontrate	Screening-intervall (Jahre)	Inkrementelle Kosten (US\$)	Inkrementelle Effekte (LT)	Inkrementelle KE-Ratio (US\$/LJ)
McCroy & Matchar 1999	USA	US\$ 1997	100%iges computerisiertes Rescreening	3%	1	436	0,7	227.343
					2	246	1,6	56.119
					3	178	2,6	24.988
Brown & Garber 1999	USA	US\$ 1996	AutoPap™ 300 QC-Rescreening	3%	1	134	0,30	163.033
					2	66	0,57	42.263
					3	43	0,96	16.349
Hutchinson et al. 2000	USA	US\$ 1997	AutoPap™ 300 QC-Rescreening	3%	1	78	0,086	331.046
					2	38	0,145	95.655
					3	25	0,189	48.280
Brown & Garber 1999	USA	US\$ 1996	PAPNET™ - Rescreening	3%	1	257	0,34	275.897
					2	129	0,63	74.738
					3	86	1,07	29.336
Radensky & Mango 1998	USA	US\$ 1997	PAPNET™ - Rescreening	3%	3	90 ^A	0,841 ^A	39.061 ^A
					3	90 ^B	0,791 ^B	41.530 ^B
					3	92 ^C	0,543 ^C	61.842 ^C

LT: gewonnene Lebenstage; LJ: gewonnene Lebensjahre.

A: Referenzstandard Biopsie; B: Referenzstandard unabhängiger Pathologie; C: Pathologie des Zentrums

II C 5. 4. 3 Einfluss der Testgüte auf die Kosten-Effektivität des Einsatzes neuer Zervixscreening-Technologien

Die in den internationalen ökonomischen Studien eingesetzten Werte für die Sensitivität und Spezifität der Screeningverfahren werden in Tabelle 45 und Tabelle 46 angegeben.

Es zeigt sich eine große Variabilität in den Sensitivitäten. Die inkrementelle Erhöhung der Sensitivität im Vergleich zur konventionellen Methode wird bei Primärscreeningverfahren zwischen 2% und 33% und bei Rescreeningverfahren zwischen 6% und 29% angegeben. Aufgrund fehlender Evidenz wurde für die Spezifität neuer Technologien häufig die Spezifität des Papanicolaou-Tests verwendet.

Die Sensitivität der verschiedenen Testverfahren ist zum einen ein bedeutender Faktor in der entscheidungsanalytischen Modellierung und unterliegt zum anderen einer relativ hohen Unsicherheit. Abbildungen 1 und 2 stellen den Zusammenhang zwischen inkrementellem Kosten-Effektivitäts-Verhältnis und inkrementeller Sensitivität der neuen Screeningtechnologien im Vergleich zum konventionellen Verfahren dar. In beiden Abbildungen ist ein eindeutiger Trend zu erkennen: Je größer die Verbesserung der Sensitivität, desto kleiner das inkrementelle Kosten-Effektivitäts-Verhältnis der neuen Technologie im Vergleich zur konventionellen Methode.

Tabelle 45: Sensitivitäten und Spezifitäten konventioneller und neuer Primärscreening-Verfahren.

Quelle	Schwellenwert	Konventioneller Papanicolaou-Test + 10% Rescreening		Primärscreening Technologie + 10% Rescreening		Referenzen
		Se (%)	Sp (%)	Se (%)	Sp (%)	
Brown & Garber 1999	LSIL+	81,6	95,8	92,6	95,8	(Sheets et al. 1995; Boon et al. 1994; Colgan et al. 1995; Jenny et al. 1997; Rosenthal et al. 1996; Eddy 1990)
Hutchinson et al. 2000	LSIL+	50,4	K. A.	55,2 ^A	K. A.	(Brown & Garber 1999)
		50,4		75,0 ^B		
McCrorry & Matchar 1999	ASCUS+	53,5	97,0	84,3	100,0	(McCrorry & Matchar 1999; Bolick & Hellman 1998; Colgan et al. 1995; Jenny et al. 1997; Patten S.F. et al. 1997; Roberts et al. 1997; Brown & Garber 1999; Stevens S. A. et al. 1997b)
Payne et al. 2000	Mild / Grenzwertig	37–100 †	98,0	52–100 †	98,0	(Sherlaw-Johnson et al. 1994)

Se: Sensitivität; Sp: Spezifität; ASCUS+: Atypical Squamous Cells of Unknown Significance; LSIL+: Low Squamous Intraepithelial Lesion; K. A. : Keine Angabe; A: Autopap; B: ThinPrep; †CIN 1 bis invasives Karzinom

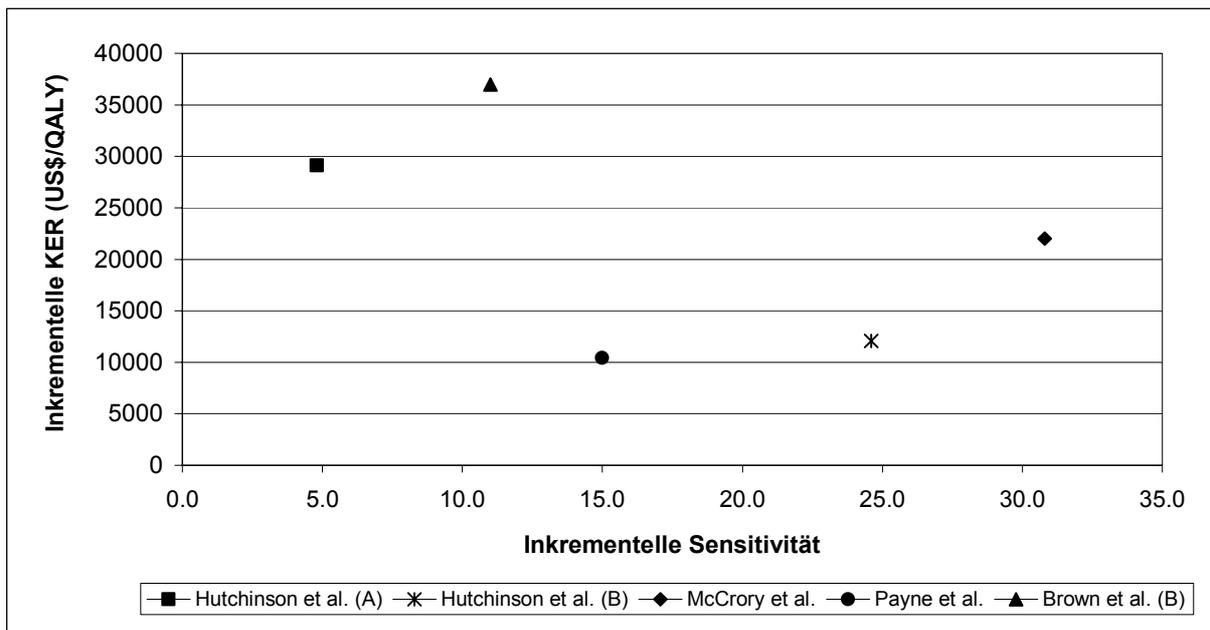


Abbildung 9: Einfluss der inkrementellen Sensitivität auf das inkrementelle Kosten-Effektivitäts-Verhältnis des Einsatzes neuer Primärscreening-Technologien versus konventioneller Methode.

KER: Kosten-Effektivitätsratio, LJ: Gewonnene Lebensjahre. A: AutoPap, B: ThinPrep. Dargestellt ist die inkrementelle Sensitivität der neuen Primärscreening-Technologie im Vergleich zu ihrem inkrementellen Kosten-Effektivitäts-Verhältnis versus der konventionellen Methode bei einem Einsatz im Screeningintervall von 3 Jahren.

Tabelle 46: Sensitivität und Spezifität manueller und automatisierter Rescreening-Verfahren.

Quelle	Schwellenwert	Papanicolaou-Test + Manuelles Rescreening		Papanicolaou-Test + Automatisiertes Rescreening		Referenzen
		Se (%)	Sp (%)	Se (%)	Sp (%)	
Brown & Garber 1999	LSIL+	81,6	95,8	95,4 ^A	95,4	(Sheets et al. 1995; Boon et al. 1994; Colgan et al. 1995; Jenny et al. 1997; Rosenthal et al. 1996; Eddy 1990)
		81,6	95,8	97,0 ^B	95,4	
Hutchinson et al. 2000	LSIL+	50,4	K. A.	55,3	K. A.	(McCrorry & Matchar 1999; Fahey et al. 1995; Brown & Garber 1999)
McCrorry & Matchar, 1999	ASCUS+	53,5	97,0	80,4	100,0	(McCrorry & Matchar 1999; Bolick & Hellman 1998; Colgan et al. 1995; Jenny et al. 1997; Patten S.F. et al. 1997; Roberts et al. 1997; Brown & Garber 1999; Stevens S. A. et al. 1997b)
Radensky & Mango 1998	ASCUS+	85,0	99,5	98,0 ^K	99,5	(Mango & Radensky 1998; Radensky & Mango 1998; Eddy 1990)

Se: Sensitivität; Sp: Spezifität; ASCUS+: Atypical Squamous Cells of Unknown Significance; LSIL+: Low Squamous Intraepithelial Lesion; K. A. : Keine Angabe; A: AutoPap; B: Papnet; K: Referenzstandard Biopsie

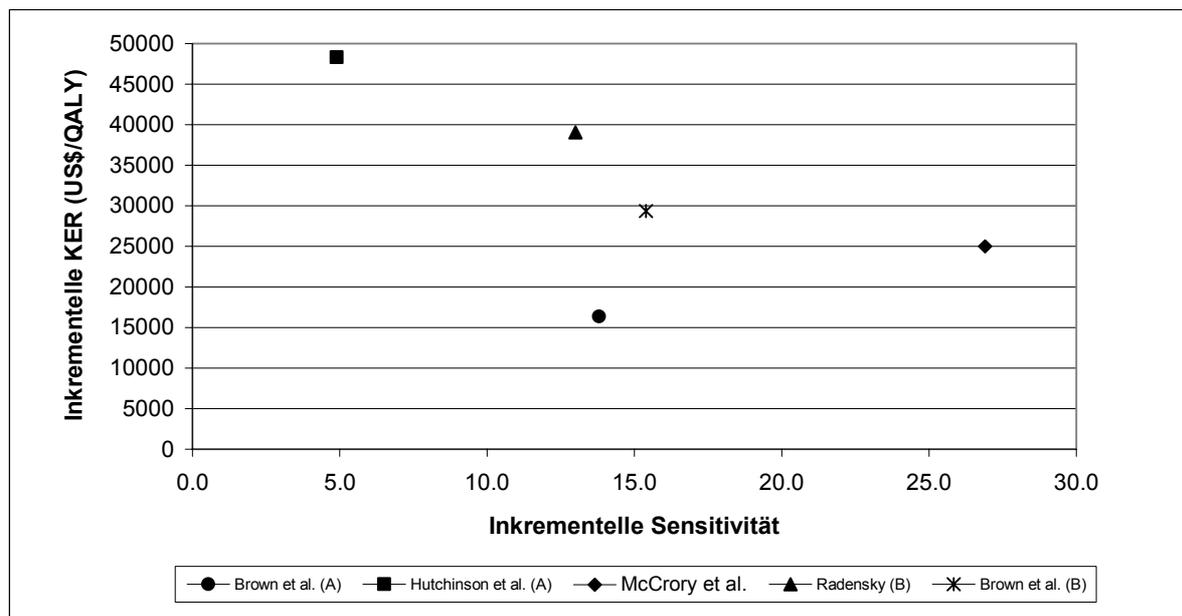


Abbildung 10: Einfluss der inkrementellen Sensitivität auf das inkrementelle Kosten-Effektivitäts-Verhältnis des Einsatzes automatisierter Rescreening-Technologien versus konventioneller Methode.

KER: Kosten-Effektivitätsratio, LJ: Gewonnene Lebensjahre, A: AutoPap, B: PapNet.

Dargestellt ist die inkrementelle Sensitivität der automatisierten Rescreening-Technologien im Vergleich zu ihrem inkrementellen Kosten-Effektivitäts-Verhältnis versus der konventionellen Methode bei einem Einsatz im Screeningintervall von 3 Jahren.

II C 6 Diskussion

Es wurden keine randomisierten kontrollierten Studien oder Feldstudien identifiziert, die den ökonomischen Aspekt eines Einsatzes neuer Screeningtechnologien (Dünnschichtpräparationen und automatisierte Auswertungsverfahren) untersucht haben. Alle berücksichtigten gesundheitsökonomischen Studien basierten auf entscheidungsanalytischen Markov-Modellen, die den Krankheitsverlauf simulieren.

Entscheidungsanalytische Markov-Modelle bieten einen systematischen, expliziten und quantitativen Ansatz, um wissenschaftliche Evidenz zu diagnostischer Testgüte und Langzeitprognose mit ökonomischen Parametern zu verknüpfen. Ferner können neben den Zielparametern Lebenserwartung und Lebenszeitkosten andere für die Entscheidungsfindung und Bedarfsplanung wichtige klinische Zielparameter evaluiert werden. Die erforderlichen Annahmen eines Modells können im Rahmen von Sensitivitätsanalysen modifiziert und ihr Einfluss auf das Analyseergebnis untersucht werden. Dieses ermöglicht außerdem eine Identifikation kritischer Parameter, für die Forschungsbedarf besteht.

Eine Vergleichbarkeit von Modellergebnissen verschiedener Studien untereinander unterliegt jedoch Limitationen unterschiedlicher Natur, die im Folgenden näher beschrieben werden.

II C 6. 1 Vergleichbarkeit und Limitationen

Evaluierte Screeningstrategien

Die Ergebnisse aus verschiedenen Studien zur Kosten-Effektivität der neuen Technologien sind bezüglich unserer Fragestellung direkt miteinander vergleichbar, wenn als Vergleichsstrategie das konventionelle Screening mit der Papanicolaou-Methode gewählt wurde. Nicht in allen evaluierten Studien war dieses der Fall. Jedoch konnten Inkremente in Bezug auf das konventionelle Verfahren als Vergleichsstrategie berechnet werden, wo in den eingeschlossenen Studien disaggregierte Ergebnisse berichtet wurden. Studien, die jedoch nicht die Strategie „konventionelles Screening“ evaluierten bzw. die disaggregierten Ergebnisse nicht berichteten, konnten nicht in der systematischen Ergebnissynthese einbezogen werden.

Analysedesign und Modellierung

Einige Studien waren auf Grund eines anderen Studiendesigns nicht mit den meisten Studien vergleichbar (Noorani 1997, AHTAC 1998). Es wurde in diesen Fällen kein entscheidungsanalytisches Markov-Modell eingesetzt, so dass keine Angaben zu Gesamtlebenszeitkosten und Überlebenszeit möglich waren. Der Zeithorizont dieser Studien betrug zwischen einem und zwei Jahren.

Perspektive

Der Standpunkt, aus dessen Sicht Kosten und Nutzen einer gesundheitsökonomischen Evaluation erfasst werden, wird Perspektive genannt. Die Perspektive einer gesundheitsökonomischen Studie bestimmt, welche Kostenarten in der Kosten-Effektivitäts-Studie berücksichtigt werden. Der umfassendste Ansatz bildet die gesamtgesellschaftliche Perspektive, die unabhängig vom Kostenträger alle Kosten berücksichtigt und von Experten als Standard für gesundheitsökonomische Studien empfohlen wird (Greiner & Hoffmann 1999; Gold et al. 1996).

In den meisten evaluierten Studien wurde die gesamtgesellschaftliche Perspektive jedoch nur approximativ bzw. gar nicht eingenommen. Einige Studien nahmen Perspektiven von Kostenträgern bzw. Gesundheitsorganisationen, des Gesundheitssystems oder des Patienten ein. Diese Perspektiven implizieren eine eingeschränkte Betrachtungsweise, die sowohl bei der Vergleichbarkeit von Studienergebnissen verschiedener gesundheitsökonomischer Studien als auch bei der Übertragbarkeit von Studienergebnissen auf andere Länder (z.B. Deutschland) berücksichtigt werden sollte.

Diskontrate

Eine Diskontierung reflektiert die generelle Präferenz des Individuums bzw. der Gesellschaft für gegenwärtige Güter im Vergleich zu in der Zukunft liegenden Gütern.

Das Washington-Panel (Gold et al. 1996) empfiehlt für den Referenzfall in gesundheitsökonomischen Studien eine Diskontierung von Kosten und Gesundheitseffekten jeweils mit derselben Diskontierungsrate. Werden die Effekte nicht diskontiert bzw. mit einer geringeren Diskontrate als die Kosten diskontiert, so kann das Kosten-Effektivitäts-Verhältnis erheblich über- oder unterschätzt werden (Drummond et al. 1997).

Raab et al. (Raab et al. 1999b) diskontierten einseitig nur die Gesundheitseffekte, welches vermutlich zu einer Unterschätzung des Kosten-Effektivitäts-Verhältnisses führte. Payne et al. (Payne et al. 2000) diskontierten in der Basisfall-Analyse Kosten mit jährlich 6% und Gesundheitseffekte mit jährlich 1,5%, welches zu einem günstigeren Kosten-Effektivitäts-Verhältnis führte. In einer Sensitivitätsanalyse wurde die Diskontierung von Kosten und Gesundheitseffekten jedoch mit jeweils jährlich 3% durchgeführt.

Ein Vergleich von Analyseergebnissen aus Studien mit unterschiedlicher Diskontierungsrate ist generell als problematisch einzuschätzen. Einen internationalen Standard gibt es nicht. Das Washington Panel empfiehlt eine Diskontierungsrate von 3% in der Basisfall-Analyse und von 5% in Sensitivitäts- bzw. Zusatzanalysen, da viele Studien in der Vergangenheit diese Diskontierungsrate einsetzten. In Deutschland wird in Sensitivitätsanalysen eine Diskontierungsrate von 0-10% empfohlen (Anonymous 1999a).

Berücksichtigte Kosten

In allen gesundheitsökonomischen Studien wurden ausschließlich direkte medizinische Kosten erhoben. Nichtmedizinische Kosten, wie Transport- und Personalkosten wurden mit Ausnahme von Raab et al. (Raab et al. 1999b) ebenso wenig berücksichtigt wie indirekte Kosten, worunter u.a. Produktivitätsausfallkosten fallen und intangible Kosten, wie Schmerz, psychische Belastung oder Lebensqualitätsverlust. McCrory et al. verwiesen als einzige Autoren auf einen dringenden Forschungsbedarf in diesem Gebiet. Insbesondere wenn die Kosten-Effektivität als Kriterium zur Ressourcenallokation über verschiedene Krankheitsbereiche im Gesundheitswesen verwendet werden soll, ist der Einbezug von Lebensqualitätsparametern wichtig, um Kosten pro qualitätskorrigiertes Lebensjahr berechnen zu können.

Die Vergleichbarkeit von Kosten zwischen Studien unterschiedlicher Länder bzw. Gesundheitssysteme ist bedingt durch die Variation der klinischen Praxis nur begrenzt möglich. Eine detaillierte Darstellung von Mengen- und Preisgerüsten, die zur Berechnung der Kosten berücksichtigt wurden, ist hierfür ebenso notwendig wie die Angabe der Währung und des Jahres der Kostenerhebung. Nicht in allen der evaluierten Studien wurden diese transparent dargestellt.

Berichtete Outcomes

Zur Vergleichbarkeit von Ergebnissen verschiedener Studien ist die Wahl von einheitlichen Outcomes und ihre transparente Darstellung unerlässlich, welches in einigen Studien nicht berücksichtigt wurde. In einigen Fällen wurden Effekte und Kosten in der Ergebnispräsentation nicht absolut pro Kopf dargestellt, welches eine Berechnung des inkrementellen Kosten-Effektivitäts-Verhältnisses nicht ermöglichte. Nicht in allen evaluierten Studien wurden Gesundheitseffekte und Kosten simultan untersucht, welches für eine sinnvolle Aussage zur Kosten-Effektivität jedoch notwendig ist. Noorani et al. (Noorani et al. 1997) berechneten lediglich die direkten Gesamtlaborkosten pro zusätzlich entdecktem Fall. Im AHTAC-Bericht (AHTAC 1998) wurden Kosten pro gefundenen abnormalen Befund in dem Screeningintervall bzw. pro potentiellen invasiven Karzinomfall berechnet. Da jedoch ein abnormaler Befund nicht unmittelbar zwingend zu einem invasiven Karzinom führt und bei einem Screening im Folgejahr entdeckt werden könnte, unterliegen diese Analyseergebnisse einer hohen Unsicherheit.

Sensitivitätsanalysen

Sensitivitätsanalysen dienen sowohl zur Evaluation des Einflusses einer Variation von wichtigen Modellvariablen auf das Studienergebnis als auch zur Identifikation von Modell-Parametern, für die weiterer Forschungsbedarf besteht.

Limitation einiger evaluierter Studien ist, dass ausschließlich univariate Sensitivitätsanalysen durchgeführt wurden, d.h. es wurde nur eine Variable zur Zeit variiert. Dieses birgt eine Annahme der Unabhängigkeit aller Modell-Parameter von einander.

Teilnahmerate

Payne et al. (Payne et al. 2000) berücksichtigten als einzige Autoren eine Teilnahmerate am Screening von insgesamt 85%. Die Teilnahmerate ist ein wichtiger Parameter, der in der gesundheitsökonomischen Evaluation berücksichtigt werden sollte, falls für das Zielland dazu Daten vorliegen. Insbesondere in Health-Policy Modellen zur entscheidungsanalytischen Evaluation von Screeningtechnologien sollte das Screening-Teilnahmeverhalten zumindest in einer Sensitivitätsanalyse berücksichtigt werden.

Interessenskonflikte

In einigen der evaluierten Studien waren Autoren nach eigenen Angaben als Berater eines Testherstellers einer neuen Screeningtechnologie tätig (Radensky & Mango 1998; Schechter 1996) oder die Studie wurde vom Hersteller finanziert (Smith et al. 1999; Hutchinson et al. 2000). Dieses ist legitim, sollte jedoch transparent gemacht werden.

II C 6. 2 Kosten-Effektivität der neuen Screeningtechnologien

Alle internationalen Kosten-Effektivitäts-Studien konnten zeigen, dass ein Einsatz der neuen Screeningtechnologien im Primär- bzw. Rescreening kosteneffektiv sein kann, wenn das Screening in Zeitabständen von drei oder mehr Jahren durchgeführt wird. Kürzere Screeningintervalle erhöhen das Kosten-Effektivitäts-Verhältnis. Die Studienergebnisse unterlagen einer breiten Streuung. Sensitivitätsanalysen haben gezeigt, dass die Modellergebnisse einer Reihe von Unsicherheiten in wichtigen Modellparametern unterliegen.

Testgüte

Die Ergebnisse bezüglich der Kosten-Effektivität der neuen Screeningtechnologien im Vergleich zum konventionellen Verfahren waren in allen Studien deutlich sensitiv für Änderungen in den Testgütekriterien, der Sensitivität und Spezifität. Diese Modellparameter unterlagen jedoch einer mangelnden Evidenzgrundlage. In Abwesenheit reliabler Daten wurden in den meisten Modellanalysen Annahmen gemacht, dass im Vergleich zur herkömmlichen Methode die Sensitivität der neuen Testverfahren höher liegt, diese Erhöhung für alle Abnormalitätsstadien gleich hoch ist, und die Spezifität unverändert bleibt. Die Schätzung der inkrementellen Sensitivitätserhöhung hatte einen Einfluss auf das Kosten-Effektivitäts-Verhältnis eines Einsatzes einer neuen Technologie im Vergleich zur konventionellen. Es konnte deskriptiv eine reziproke Korrelation zwischen der relativen inkrementellen Sensitivität und der Höhe des inkrementellen Kosten-Effektivitäts-Verhältnisses gezeigt werden. Insbesondere die Höhe der Sensitivität des konventionellen Papanicolaou-Tests als Basis zur Bestimmung der inkrementellen Sensitivitätserhöhung variierte in den verschiedenen Studien beträchtlich. Eine reliable und vergleichbare Schätzung der Sensitivität und Spezifität ist jedoch Voraussetzung für eine aussagekräftige Kosten-Effektivitätsanalyse. Die Annahme einer gleich hohen Verbesserung der Sensitivität unabhängig vom Abnormalitätsgrad ist nicht valide (Broadstock 2000). Überschätzte inkrementelle Sensitivitätserhöhungen für hochgradige Abnormalitäten können zu

verfälschten Effektivitätsschätzungen für die neuen Technologien führen. Eine Erhöhung der Sensitivität kann jedoch auch mit einer Reduktion der Spezifität einhergehen. Die Erhöhung der falsch-positiven Befundrate kann einen erheblichen Einfluss auf Kosten und auf die Lebensqualität haben, welches zu einer drastisch geringeren Kosten-Effektivität führen kann.

Es besteht deshalb Forschungsbedarf in der Erhebung valider Schätzungen für die Sensitivität und Spezifität der Screeningtechnologien.

Testkosten

Die Ergebnisse der Modellanalysen unterlagen weiterhin gegenüber Änderungen in den Kosten der neuen Technologien einer breiten Streuung. Test-Kosten basierten häufig ausschließlich auf Verbrauchsmaterial- und Kapitalkosten. Kosten für den Geräteservice, Raumnutzung, Mitarbeitertraining, Durchführung der Tests, Transport und Lagerung blieben häufig unberücksichtigt, so dass eine Unterschätzung der tatsächlichen Testkosten anzunehmen ist. Kosten für administrative Tätigkeiten können bei den neuen computergestützten Verfahren durchaus höher sein als bei der konventionellen Methode. Kosten für die Testdurchführung sind abhängig von der benötigten Zeit pro Testung und sind aus diesem Grund oftmals schwierig objektiv zu bestimmen. Für große Labore mit großer Kapazität können sich jedoch die Kosten pro Einheit reduzieren.

Screeningintervall

Je kürzer das Screeningintervall, um so teurer ist der Einsatz eines Screeningverfahrens. Dieses ist in allen gesundheitsökonomischen Studien ersichtlich. Ein Einsatz der neuen Technologien in einem Screeningintervall von weniger als drei Jahren ist mit einem erheblich höheren Kosten-Effektivitäts-Verhältnis verbunden und nicht mehr als kosteneffektiv einzustufen. Die Ursachen hierfür sind vergleichsweise häufiger durchgeführte Screeningtests und somit mehr erkannte Dysplasien, die behandelt werden „müssen“. Bei den erkannten Krebsvorstufen handelt es sich jedoch meistens um geringgradige Veränderungen, die sich ohne Behandlung zum größten Teil spontan zurückbilden oder bei Persistenz auch bei einem längeren beispielsweise 3-jährigen Screeningintervall noch rechtzeitig erkannt und behandelt werden können, da beim Zervixkarzinom bekanntlich langsame Progressionszeiten vorliegen (Sawaya et al. 2001).

Unter der Prämisse des jährlichen Screenings wird ein Einsatz der neuen Technologien in der Zervixkarzinomfrüherkennung in Deutschland nicht als kosteneffektiv einzustufen sein. Eine Verlängerung des Intervalls auf ein dreijährliches Screening bei Einsatz neuer Technologien würde jedoch zur Diskussion stehen. Letztendlich wird die Länge der Screeningintervalle von der Zahlungsbereitschaft (willingness to pay) der Gesellschaft abhängen.

Lebensqualität

Qualitätskorrigierte Lebensjahre (QALY) sind bevorzugtes Effektivitätsmaß einer Kosten-Effektivitäts-Studie. Aus Mangel an evidenten Daten berücksichtigte jedoch keine der Studien die Lebensqualität, obgleich angenommen wird, dass ein positives Testergebnis psychologisch-traumatische Auswirkungen auf einige Frauen haben kann (Bell 1995). Angst vor bevorstehenden Untersuchungen, Diagnosen und Behandlungen können die Ursache sein. Falsch-positive Testergebnisse können aus diesem Grund unter Berücksichtigung der Lebensqualität in Kosten-Effektivitäts-Analysen eine erhebliche Auswirkung auf das inkrementelle Kosten-Effektivitäts-Verhältnis haben.

Zukünftiger Forschungsbedarf liegt in der Erhebung valider Lebensqualitätsdaten von Frauen mit verschiedenen Screeningtestergebnissen die sich in unterschiedlichen Gesundheitsstadien der Zervixkarzinomkrankung befinden (HPV-Infektion, Dysplasiestufen, Karzinomstufen).

II C 6. 3 Übertragbarkeit der Ergebnisse auf den Kontext des deutschen Versorgungssystems

Für Deutschland liegt bislang keine gesundheitsökonomische Studie zum Zervixkarzinom-Screening vor. Die evaluierten gesundheitsökonomischen Studien stammen aus Australien, Großbritannien, Kanada und den USA. Ob und in wieweit eine Übertragbarkeit dieser Studienergebnisse auf das Gesundheitssystem in Deutschland möglich ist, hängt von vielen Faktoren ab.

Nach O'Brien (O'Brien 1997) sowie Greiner et al. (Greiner & Uber 2000) ist eine Übertragung von Studienergebnissen auf andere Länder nur dann möglich, wenn folgende Kriterien übereinstimmen: Demographie und Epidemiologie der Erkrankung, klinische Praxis und Konventionen, Anreizsysteme und Regulationsmechanismen für Leistungserbringer im Gesundheitssystem, relative Preise, Verbraucherpräferenzen

und Opportunitätskosten. Welte und Leidl (Welte & Leidl 1998) ergänzen diesen Katalog um Elemente aus dem Kosten-und-Effekte-Sektor (z.B. Perspektive, Diskontrate), dem technologischen Kontext (z.B. personelle Charakteristika des medizinischen Personals bzw. von Patienten), sowie dem epidemiologischen Kontext (z.B. Fallmischung). Diese Vielfalt abhängiger Faktoren zeigt deutlich, welche komplexe Aufgabe eine Übertragung ausländischer Studienergebnisse auf deutsche Verhältnisse darstellt.

Welte und Leidl (Welte & Leidl 1998) bieten in ihrem Artikel auch Lösungsmöglichkeiten zur Verbesserung der Übertragbarkeit von Studienresultaten an. Durch Anpassung des Mengen- und Preisgerüsts kann die Übertragbarkeit beispielsweise verbessert werden. Dies erscheint jedoch insofern schwierig, als nur in den wenigsten Fällen Mengen- und Preisgerüste im Detail angegeben werden. Weiterhin werden die Anpassung an die Zeit (mittels OECD-Indizes) und an das jeweilige Währungssystem (z.B. Kaufkraftparitäten, PPP) genannt (Welte & Leidl 1998).

Aufgrund der erheblichen Heterogenitäten der einzelnen Studien ist eine Übertragbarkeit auf deutsche Verhältnisse nicht realistisch. Die notwendigen Anpassungsverfahren sind sehr aufwendig und in vielen Bereichen aufgrund mangelnder Datenlage nicht möglich. Als Alternative ist die Entwicklung eines für Deutschland spezifischen entscheidungsanalytischen Modells zu sehen. Ein systematisches Vorgehen zur Übertragung von Evidenz aus Studien in anderen Ländern auf den Kontext des deutschen Gesundheitssystems wurde in einem HTA-Methodenreport im Rahmen der HTA-Projektreihe der Deutschen Agentur für Health Technology Assessment beim DIMDI entwickelt (Siebert U. 2002b; Siebert U 2002a).

II C 7 Schlussfolgerungen

- Basierend auf internationalen Kosten-Effektivitäts-Studien kann der Einsatz neuer Zervixkarzinom-Screening-Technologien (Dünnschichtpräparation und computergestützte Auswertungsverfahren) zu akzeptablen Kosten pro gewonnenem Lebensjahr führen, wenn das Screening in einem Zeitintervall von drei oder mehr Jahren erfolgt.
- Die bestehenden Unsicherheiten im Bereich der Effektivität und Kosten der neuen Verfahren erschweren jedoch eine valide Aussage zur Bestimmung einer optimalen Screening-Strategie und eines optimalen Screeningintervalls. Forschungsbedarf besteht in validen Erhebungen der Testgüte neuer Verfahren unter Verwendung von Goldstandards bei der durchschnittlichen Bevölkerung.
- In Anbetracht der in Deutschland festgelegten jährlichen Screeningpraxis kann die Einführung der Dünnschichtpräparation und/oder computergestützter Auswertungsverfahren auf Basis der publizierten und bewerteten internationalen Evidenz nicht als kosteneffektiv bewertet werden.
- Aufgrund der internationalen Daten ist auch für Deutschland eine Verlängerung des Screeningintervalls bei eventuellem routinemäßigem Einsatz neuer Screeningtechnologien zu diskutieren.
- Es besteht Handlungs- und Forschungsbedarf bezüglich der Verbesserung der Teilnahme der Frauen am regelmäßigen Screening, insbesondere in unterversorgten Subpopulationen mit mangelhafter Screeningteilnahme.
- Die Entwicklung eines entscheidungsanalytischen Modells für den deutschen Versorgungskontext war notwendig, um die Langzeit-Effektivität und Kosten-Effektivität neuer Untersuchungsmethoden von Zervixabstrichen im Kontext des deutschen Gesundheitswesens evaluieren zu können.

II C 8 Literatur

Die Referenzen zu der in Kapitel II und III zitierten Literatur einschließlich der ein- und ausgeschlossenen Studien sind gemeinsam im Abschnitt III am Ende des Berichts aufgeführt.

II Anhang

II 1 Dokumente der Literatursuche

Übersicht Datenbanken

Zeitraumen: 02.01.2001 bis 07.05.2002

Datenbanken	Jahrgänge/Monat (Woche)
Best Evidence	von 1991 bis 2002
CancerLit	von 1975 bis 2001
CINAHL	von 1982 bis 2001
Cochrane Library	2000/10 bis 12
DARE	nicht bekannt
EconLit	Von 1969 bis 2002
Embase	von 1988 bis 2001 (5)
HealthStar	von 1975 bis 2001
Medline Express	von 1986 bis 2001
Premedline/Medline	von 1966 bis aktuell
NEED	nicht bekannt
NHS HTA-Database	nicht bekannt
PubMed	von 1966 bis 2001
Web of Science (Science Citation Index, Social Science Citation Index and Arts & Humanities Citation Index)	von 1989 bis 2001

Datenbank-Suchstrategien für gesundheitsökonomische Evaluation

Datenbank: Cochrane Library
 Software: Internet
 Datum der Recherche: 02.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	automated cervical cytology + vaginal smears	

Datenbank: NHS Centre for Reviews & Dissemination (DARE, NHS EED & HTA databases)
 Software: Internet
 Datum der Recherche: 02.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	cervical cytology + vaginal smears	57

Datenbank: PubMed
 Datum der Recherche: 02.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	automated cervical screening	5

Datenbank: UK NCCHTA
 Software: Internet
 Datum der Recherche: 02.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	automated cervical screening	1

Datenbank: Minnesota Health Technology Advisory Committee (HTAC)
 Software: Internet
 Datum der Recherche: 02.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	automated cervical screening	1

Datenbank: Canadian Coordinating Office for Health Technology Assessment (CCOHTA)
 Software: Internet
 Datum der Recherche: 02.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	automated cervical screening	1

Datenbank: PubMed (1966 bis 2001)
 Limitiert ab 1985, human studies

Software: Ovid
Datum der Recherche: 06.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	cervix neoplasms [mh] OR cervical intraepithelial neoplasia [mh] OR cervix dysplasia [mh] OR papillomavirus, human [mh] AND vaginal smears [mh] OR papanicolaou [ti] OR PAP [ti] OR cytological techniques [mh] OR histocytological preparation techniques [ti] OR cytodiagnosis [mh] OR diagnosis [sh] AND PapNet OR AutoPap OR ThinPrep OR Autocyte OR Cytorich OR "Cyto Rich" OR "Easy Prep" OR thinlayer OR monolayer AND sensitivity and specificity OR sensitivity [tw] OR specificity [tw] OR screen* [ti] OR rescreen* [ti] AND human	115
	#2	PapNet OR AutoPap OR ThinPrep OR Autocyte OR Cytorich OR "Cyto Rich" OR "Easy Prep"	245

Datenbank: EMBASE (1988 bis Oktober 2000)
Software: Ovid
Datum der Recherche: 06.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	exp *uterine cervix cancer/di OR exp *uterine cervix carcinoma in situ/di OR exp *uterine cervix dysplasia/di OR exp *wart virus/ AND exp *vagina smear/ OR exp *cancer screening/ OR exp *papanicolaou test/ OR exp *uterine cervix cytology/ OR exp *cytology/ OR exp *histology/ OR exp *cancer cytodagnosis/ OR exp *cytodagnosis/ OR exp *mass screening/ OR exp *screening/ OR exp *screening test/ OR "screening".mp. OR "rescreen\$".mp. AND "sensitivity".mp. OR "specificity".mp.	200
	#2	(PapNet OR AutoPap OR ThinPrep OR Autocyte OR Cytorich OR "Cyto Rich" OR "Easy Prep").mp. [mp=title, abstract, heading word, trade name, manufacturer name]	213

Datenbank: CINAHL (1982 bis Oktober 2000)
 Software: Ovid
 Datum der Recherche: 06.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	exp *cervix neoplasms/di OR "cervical intraepithelial neoplasia".mp. OR "cervix dysplasia".mp. OR "uterine cervix carcinoma in situ".mp. OR exp *papillomaviruses/ OR "human papillomavirus".mp. AND exp *cervical smears/ OR "pap smear".mp. OR "papanicolaou".mp. OR exp *cancer screening/ OR exp *cytodiagnosis/ OR exp *cytological techniques/ OR "cytology".mp. OR exp *histology/ OR "histology".mp. OR "mass screening".mp. OR "screening".mp. OR "re-screen\$".mp. AND exp *"sensitivity and specificity"/ OR "sensitivity and specificity".mp.	5
	#2	OR (PapNet OR AutoPap OR ThinPrep OR Autocyte OR Cytorich OR "Cyto Rich" OR "Easy Prep").mp. [mp=title, cinahl subject heading, abstract, instrumentation]	14

Datenbank: HealthSTAR (1975 bis 2000)
 Medline Referenzen ausgeschlossen
 Software: Ovid
 Datum der Recherche: 06.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	cervix neoplasms OR cervix dysplasia OR cervical intraepithelial neoplasia OR papillomavirus, human AND screening AND sensitivity and specificity	14
	#2	AutoPap OR PapNet OR ThinPrep OR Autocyte OR Cytorich OR "Cyto Rich" OR "Easy Prep"	4

Datenbank: CancerLit (1983 bis November 2000)
 Software: Ovid
 Datum der Recherche: 06.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	exp *cervix neoplasms/di OR "cervical intraepithelial neoplasia".mp. OR exp *cervix dysplasia/di OR exp *papillomavirus/ AND exp *vaginal smears/ OR "papanicolaou".mp. OR "pap smear".mp. OR exp *cytological techniques/ OR exp *cytodiagnosis/ OR "cytodiagnosis".mp. OR exp *diagnosis/ OR exp *diagnosis, computer-assisted/ AND exp **sensitivity and specificity"/ OR "sensitivity and specificity".mp. Jahrgang 1983-1992	34
	#2	Jahrgang 1993-1996	87
	#3	Jahrgang 1997-November 2000	173
	#4	(PapNet OR AutoPap OR ThinPrep OR Autocyte OR Cytorich OR "Cyto Rich" OR "Easy Prep").mp. [mp=title, abstract, mesh subject heading] Jahrgang 1983-1992	1
	#5	Jahrgang 1993-1996	36
	#6	Jahrgang 1997- November 2000	135

Datenbank: Web of Science (Science Citation Index, Social Science Citation Index & Arts & Humanities Citation Index) (1989 bis 2001)
 Software: Internet
 Datum der Recherche: 06.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	(PapNet OR AutoPap OR ThinPrep OR Autocyte OR Cytorich OR "Cyto Rich" OR "Easy Prep" OR "automated cervical cancer") AND (cervical OR cervix OR papillomavirus)	211

Datenbank Best Evidence (1991 bis 2001)
 Software: Internet
 Datum der Recherche: 06.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	cervical cancer.mp. [mp=title, abstract, full text, keywords, caption text]	7

Datenbank: PubMed (1966 bis 2001)
 Limitiert ab 1999
 Software: Ovid
 Datum der Recherche: 05.03.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	cervix neoplasms OR cervical intraepithelial neoplasia OR cervix dysplasia OR vaginal smears OR cytological techniques OR histocytological preparation techniques OR cytodiagnosis AND "fluid based" OR thinlayer OR thinprep OR "thin prep" OR "thin layer" OR monolayer* OR "mono layer*" OR liquid* OR cytyc OR cytorich OR "cytorich" OR autocyte OR autopap OR papnet AND sensitivity and specificity OR sensitivity OR specificity OR diagnosis OR pathology	669

Datenbank: Embase (1988 bis 2001)
 Limitiert ab 1999
 Software: Ovid
 Datum der Recherche: 05.03.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	uterine cervix tumor / OR cervix neoplasms.mp. OR uterine cervix cancer/ OR cervical cancer.mp. OR uterine cervix carcinoma in situ/ OR cervical intraepithelial neoplasia.mp. OR vagina smear/ OR vaginal smears.mp. OR cytology/ OR cytological techniques.mp. OR histology/ OR histocytological preparation techniques.mp. OR cancer cytodiagnosis/ OR cytodiagnosis/ or cytodiagnosis.mp. AND fluid based.tw. OR thinlayer.tw. OR thinprep.tw. OR (thin adj3 prep\$.tw. OR (thin adj3 layer\$.tw. OR monolayer\$.tw. OR (mono adj3 layer\$.tw. OR liquid\$.tw. OR cytyc.tw. OR cytorich.tw. OR cytorich.tw. OR autocyte prep.tw. OR autopap.tw. OR papnet.tw. AND exp "sensitivity and specificity"/ OR sensitivity.tw. OR exp diagnosis/ OR exp pathology/ OR specificity.tw.	106

Datenbank: CINAHL (1982 bis 2001)

Limitiert ab 1999

Software: Ovid

Datum der Recherche: 05.03.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	uterine cervix tumor/ OR cervix neoplasms.mp. OR uterine cervix cancer/ OR cervical cancer.mp. OR uterine cervix carcinoma in situ/ OR cervical intraepithelial neoplasia.mp. OR vagina smear/ OR vaginal smears.mp. OR cytology/ OR cytological techniques.mp. OR histology/ OR histocytological preparation techniques.mp. OR cancer cytodagnosis/ OR cytodagnosis/ or cytodagnosis.mp. AND fluid based.tw. OR thinlayer.tw. OR thinprep.tw. OR (thin adj3 prep\$.tw. OR (thin adj3 layer\$.tw. OR monolayer\$.tw. OR (mono adj3 layer\$.tw. OR liquid\$.tw. OR cytyc.tw. OR cytorich.tw. OR cyto rich.tw. OR autocyte prep.tw. OR autopap.tw. OR papnet.tw. AND exp "sensitivity and specificity"/ OR sensitivity.tw. OR exp diagnosis/ OR exp pathology/ OR specificity.tw.	8

Datenbank: Cancerlit (1975 bis 2000)

Limitiert ab 1999

Software: Ovid

Datum der Recherche: 05.03.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	uterine cervix tumor/ OR cervix neoplasms.mp. OR uterine cervix cancer/ OR cervical cancer.mp. OR uterine cervix carcinoma in situ/ OR cervical intraepithelial neoplasia.mp. OR vagina smear/ OR vaginal smears.mp. OR cytology/ OR cytological techniques.mp. OR histology/ OR histocytological preparation techniques.mp. OR cancer cytodagnosis/ OR cytodagnosis/ or cytodagnosis.mp. AND fluid based.tw. OR thinlayer.tw. OR thinprep.tw. OR (thin adj3 prep\$.tw. OR (thin adj3 layer\$.tw. OR monolayer\$.tw. OR (mono adj3 layer\$.tw. OR liquid\$.tw. OR cytyc.tw. OR cytorich.tw. OR cyto rich.tw. OR autocyte prep.tw. OR autopap.tw. OR papnet.tw. AND exp "sensitivity and specificity"/ OR sensitivity.tw. OR exp diagnosis/ OR exp pathology/ OR specificity.tw.	78

Datenbank: Web of Science (Science Citation Index, Social Science Citation Index, and Arts & Humanities Citation Index) (1989 bis 2001)

Limitiert ab 1999

Software: Ovid

Datum der Recherche: 05.03.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	(cervix neoplasms OR cervical intraepithelial neoplasia OR cervix dysplasia OR vaginal smears OR histocytological preparation techniques OR cytodiagnosis) AND (thinlayer OR monolayer OR cytc* OR cytorich OR autocyte OR autopap OR papnet) AND (sensitivity OR specificity OR diagnosis OR pathology)	40

Datenbank: HealthSTAR (1993 bis 2000)

Limitiert ab 1999, Medline Referenzen ausgeschlossen

Software: Ovid

Datum der Recherche: 05.03.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	cervix neoplasms OR vaginal smears OR cervix dysplasia OR cervical intraepithelial neoplasia AND sensitivity and specificity	2

Datenbank: Medline Express (1986 bis Dezember 2000), Embase (1989-2000/12)

Software: Silverplatter

Datum der Recherche: 02.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	(cervix or cervical cancer) and screening and (German in la)	245
	#2	papanicolaou and (German in la)	104
	#3	(papanicolaou or test) and (German in la)	26631
	#4	(cervix or cervical cancer) and (German in la)	4457
	#5	#3 AND #4	196
	#6	(Thinprep or Papnet or autopap) and (German in la)	0
	#7	automated test and (German in la)	6

Datenbank: Medline, EBM, EconLit
 Limitiert von 2000 bis 2002
 Software: Ovid
 Datum der Recherche: 25.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
B	#1	(cervix and cancer).mp. [mp=ti, ot, ab, rw, sh, tx, kw, ct, hw]	6580
		(liquid-based or fluid or ThinPrep or Autocyte-Prep or monolayer).mp. [mp=ti, ot, ab, rw, sh, tx, kw, ct, hw]	170804
		(Papnet or AutoPap or Cytoc or Autocyte).mp. [mp=ti, ot, ab, rw, sh, tx, kw, ct, hw]	189
		(automated analyses or computer assisted analyses or neural network assisted analyses).mp. [mp=ti, ot, ab, rw, sh, tx, kw, ct, hw]	107
		(cost or economic or cost-effective).mp. [mp=ti, ot, ab, rw, sh, tx, kw, ct, hw]	325421
		(2 or 3 or 4) and 1 and 5	5
		limit 6 to yr=2000-2002	4
		2 and 1 and 5	5
		limit 8 to yr=2000-2002	4
		1 and 3 and 5	1
		1 and 4 and 5	0

Datenbank: PubMed (1966 bis 2000)
 Limitiert ab 1999
 Software: Ovid
 Datum der Recherche: 25.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
B	#1	vaginal smears OR cervix neoplasms OR cervical intraepithelial neoplasia OR cervix dysplasia AND diagnosis OR di[sh] AND "fluid based" OR thinlayer OR thinprep OR "thin prep" OR monolayer* OR "mono layer" OR liquid* OR cytoc* OR cytorich OR "cyto rich" OR "autocyte prep" OR papnet AND economics OR costs and cost analysis OR economic value of life OR economics, hospital OR economics, medical OR economics, nursing OR models, economic OR economics, pharmaceutical OR "fees and charges" OR budgets OR cost OR costs OR costed OR costly OR costing OR economic* OR pharmacoeconomic* OR price* OR pricing	45

Datenbank: HealthSTAR (1993 bis 2001)
 Limitiert ab 1999, Medline Referenzen ausgeschlossen
 Software: Ovid
 Datum der Recherche: 25.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
B	#1	vaginal smears OR cervix neoplasms OR cervical intraepithelial neoplasia OR cervix dysplasia AND economics OR costs OR costing OR cost OR budgets	42
	#2	autopap OR papnet OR autocyte OR cytorich OR thinprep AND economics OR costs OR costing OR cost OR budgets	3

Datenbank: CancerLit (1975 bis November 2000)
 Limitiert ab 1999, Medline Referenzen ausgeschlossen
 Software: Ovid
 Datum der Recherche: 25.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
B	#1	vaginal smears/ OR cervix neoplasms/ OR cervical intraepithelial neoplasia/ OR cervix dysplasia/ AND di.fs. OR exp diagnosis/ AND (fluid based OR thinlayer OR thinprep).tw. OR (thin adj3 prep\$) OR (thin adj3 layer\$).tw. OR (monolayer\$ OR mono adj3layer\$ OR liquid\$ OR cytc).tw. OR (cytorich OR cyto rich OR autocyte prep OR papnet OR autopap).tw. AND economics/ OR exp cost/ OR exp economics/	0

Datenbank EMBASE (1988 bis 2001, 7. Woche)
 Limitiert ab 1999, Medline Referenzen ausgeschlossen
 Software: Ovid
 Datum der Recherche: 25.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
B	#1	vaginal smears/ OR cervix neoplasms/ OR cervical intraepithelial neoplasia/ OR cervix dysplasia/ AND di.fs. OR exp diagnosis/ AND (fluid based OR thinlayer OR thinprep).tw. OR (thin adj3 prep\$) OR (thin adj3 layer\$).tw. OR (monolayer\$ OR mono adj3layer\$ OR liquid\$ OR cytyc).tw. OR (cytorich OR cyto rich OR autocyte prep OR papnet OR autopap).tw. AND economics/ OR exp cost/ OR exp economics/	2

Datenbank: Medline Express (1986 bis Dezember 2000), Embase (1989 bis Dezember 2000)
 Software: Silverplatter
 Datum der Recherche: 02.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
B	#1	(cervical cancer or cervix) and cost* and (Ger- man in la)	55

Datenbank PubMed (1966 bis 2001)
 Software: Ovid
 Datum der Recherche: 22.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
C	#1	cervix neoplasms [majr] OR cervical intraepithe- lial neoplasia [majr] AND epidemiology [ti] OR epidemiology [majr] OR etiology [ti] OR etiology [majr] OR "natural history" publication type: review human studies	279

Datenbank EMBASE (1988 bis Oktober 2000)

Software: Ovid

Datum der Recherche: 22.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
C	#1	exp *uterine cervix cancer/et,ep [Etiology, Epidemiology] AND exp history/ or "natural history".mp. human studies	62

Datenbank CINAHL

Software: Ovid

Datum der Recherche: 22.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
C	#1	exp *cervix neoplasms/cl,pr,ep,et [Classification, Prognosis, Epidemiology, Etiology] OR exp *cervix neoplasms/ AND "natural history".mp.	28

Datenbank CANCERLIT (1993 bis 2001)

Software: Ovid

Datum der Recherche: 22.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
C	#1	cervix neoplasms [MeSH Word (majr)] AND "natural history" OR etiology [all fields] Publication type: Review human studies only	284

Datenbank: Medline Express (1986 bis Dezember 2000), Embase (1989 bis 2000)

Software: Silverplatter

Datum der Recherche: 02.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	(cervix or cervical cancer) and (German in la)	4457
	#2	epidemiolog* and (German in la)	26699
	#3	#1 AND #2	357
	#4	(treatment or therapy) and (German in la)	220790
	#5	#3 AND #4	196

Datenbank PubMed (1966 bis 2001)
Software: Ovid
Datum der Recherche: 22.01.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
D	#1	cervix neoplasms [majr] OR cervical intraepithelial neoplasia [majr] AND "natural history" OR epidemiology [sh] OR etiology [sh] AND model* [ti] human studies	50

Datenbank PubMed (1966 bis 2001)
Software: Ovid
Datum der Recherche: 13.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
D	#1	vaginal smears OR cervix neoplasms AND (mass screening OR cytodiagnosis) AND models, theoretical OR models, organizational OR models, statistical OR markov chains	227

Datenbank: EMBASE (1988 bis 2001, 5. Woche)
Software: Ovid
Datum der Recherche: 13.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
D	#1	exp vagina smear / OR exp Papanicolaou test/ OR exp uterine cervix tumor/ OR exp uterine cervix cancer/ AND exp cancer cytodiagnosis/ OR exp cytodiagnosis/ OR exp mass screening/ AND exp model/ OR exp cancer model/ OR exp computer model/ OR exp disease model/ OR exp population model/ OR exp statistical model/ OR exp stochastic model/ OR exp mathematical model/ OR exp probability/	63

Datenbank: Cochrane Database of Systematic Reviews (2000, 4. Quartal)

Software: Internet

Datum der Recherche: 13.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
D	#1	(vaginal smears OR cervix neoplasms OR cytodiagnosis OR cervical cancer).mp. [mp=title, short title, abstract, full text, keywords, caption text] AND mass screening.mp.	1

Datenbank: Web of Science (Science Citation Index, Social Science Citation Index and Arts & Humanities Citation Index) (1989 bis 2001)

Software: Internet

Datum der Recherche: 13.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
D	#1	(vaginal smears OR cervix neoplasms OR cervical cancer OR cytodiagnosis) AND screening AND (models OR modelling OR markov)	62

Datenbank: HealthSTAR (1975 bis 2000)

Software: Ovid

Datum der Recherche: 13.02.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
D	#1	vaginal smears OR cervix neoplasms OR cytodiagnosis AND mass screening AND models OR markov chains	7

Datenbank: Premedline, Medline (1966 bis 2001)
 Software: Ovid
 Datum der Recherche: 25.03.2001 bis 09.04.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	cervical cancer and german\$	68
	#2	#1 and test	8
	#3	#1 and histology\$	5
	#4	#1 and cytol\$	11
	#5	#1 and screening	29
	#6	#1 and mass screen\$	20
	#7	#1 and rescreen\$	0
	#8	#1 and sens\$	1
	#9	#1 and spe\$	8
	#10	#1 and Pap\$	19
	#11	#1 and autom\$	0
	#12	#1 and ThinPrep	0
	#13	#1 and Autopap	0
	#14	#1 and CytoRich	0
	#15	#1 and monolayer	0
	#16	#1 and thinlayer	0
	#17	#1 and liquid\$	0
B	#1	cervix or cervical cancer and german\$	211
	#2	#1 and cost\$	10
	#3	#1 and effect\$	46
	#4	#1 and charge\$	1
	#5	#1 and fee\$	3
	#6	#1 and budget	1
C	#1	cervix or cervical cancer and german\$	211
	#2	#1 and etiolog\$	1
	#3	#1 and diag\$	65
	#4	#1 and therapy	39
	#5	#1 and treatment	43
	#6	#1 and epidemiol\$	26
	#7	#1 and preval\$	18
	#8	#1 and incidence	46
	#9	#1 and mortality	26
D	#1	cervix or cervical cancer and german\$	211
	#2	#1 and model\$	13
	#3	#1 and markov	0
	#4	#1 and decision and anal\$	1

Datenbank: Cancerlit (1975 bis November 2000)
 Software: Ovid
 Datum der Recherche: 25.03.2001 bis 09.04.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	cervical cancer and german\$	67
	#2	#1 and test	5
	#3	#1 and histology\$	6
	#4	#1 and cytol\$	15
	#5	#1 and screening	30
	#6	#1 and mass screen\$	20
	#7	#1 and rescreen\$	1
	#8	#1 and sens\$	0
	#9	#1 and spe\$	17
	#10	#1 and Pap\$	17
	#11	#1 and autom\$	0
	#12	#1 and ThinPrep	0
	#13	#1 and Autopap	0
	#14	#1 and CytoRich	0
	#15	#1 and monolayer	0
	#16	#1 and thinlayer	0
	#17	#1 and liquid\$	0
B	#1	cervix or cervical cancer and german\$	67
	#2	#1 and cost\$	2
	#3	#1 and effect\$	16
	#4	#1 and pric\$	2
	#5	#1 and fee\$	0
	#6	#1 and budget	1
C	#1	cervix or cervical cancer and german\$	67
	#2	#1 and etiolog\$	2
	#3	#1 and therapy	32
	#4	#1 and treatment	31
	#5	#1 and epidemiol\$	29
	#6	#1 and preval\$	9
	#7	#1 and incidence	23
	#8	#1 and mortality	32
D	#1	cervix or cervical cancer and german\$	67
	#2	#1 and model\$	8
	#3	#1 and markov	0

Datenbank: HealthStar (1975 bis Dezember 2000)

Software: Ovid

Datum der Recherche: 25.03.2001 bis 09.04.2001

Gruppe:	Rechercheschritt:	Suchbegriff:	Treffer:
A	#1	cervix or cervical cancer and german\$	54
	#2	#1 and test	5
	#3	#1 and histology\$	5
	#4	#1 and cytol\$	9
	#5	#1 and screening	23
	#6	#1 and mass screen\$	20
	#7	#1 and rescreen\$	0
	#8	#1 and sens\$	1
	#9	#1 and spe\$	12
	#10	#1 and Pap\$	12
	#11	#1 and autom\$	0
	#12	#1 and ThinPrep	0
	#13	#1 and Autopap	0
	#14	#1 and CytoRich	0
	#15	#1 and monolayer	0
	#16	#1 and thinlayer	0
	#17	#1 and liquid\$	0
B	#1	cervix or cervical cancer and german\$	54
	#2	#1 and cost\$	4
	#3	#1 and effect\$	11
	#4	#1 and charge\$	0
	#5	#1 and fee\$	0
	#6	#1 and budget	0
C	#1	cervix or cervical cancer and german\$	54
	#2	#1 and etiolog\$	1
	#3	#1 and therapy	12
	#4	#1 and treatment	12
	#5	#1 and epidemiol\$	9
	#6	#1 and preval\$	3
	#7	#1 and incidence	19
	#8	#1 and mortality	13
D	#1	cervix or cervical cancer and german\$	54
	#2	#1 and model\$	6
	#3	#1 and markov	0

II 2 Checkliste zur Beurteilung der methodischen Qualität gesundheitsökonomischer Studien

Siehe die folgenden Seiten.

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: Schechter 1996	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		1
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		1
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		1
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		1
6. Wurde die Zielpopulation klar beschrieben?		1
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		1
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		1
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		1
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		1
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		1
13. Wurden die relevanten Annahmen explizit formuliert?		½
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		1
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		1
16. Wurden adäquate Quellen für die Gesundheitseffektdata gewählt und eindeutig genannt?		½
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		½
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		n.r.
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		1
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		½
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		1
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		½
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		½
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	1
29. Wurden Währungskonversionen adäquat durchgeführt?	n.r.
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	1
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	1
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	1
33. Wurden die Diskontraten angegeben?	1
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	½
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	1
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	0
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	1
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	1
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	0
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	1
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	0
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	1
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	1
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	½
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	1
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	½
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	½
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	0
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	½
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	0
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	½
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	1
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	½
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	1

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: Noorani et al. 1997	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		½
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		1
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		1
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		1
6. Wurde die Zielpopulation klar beschrieben?		0
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		½
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		1
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		1
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		½
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		1
13. Wurden die relevanten Annahmen explizit formuliert?		½
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		½
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		½
16. Wurden adäquate Quellen für die Gesundheitseffektdaten gewählt und eindeutig genannt?		½
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		n.r.
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		1
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		½
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		½
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		½
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		½
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		½
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	½
29. Wurden Währungskonversionen adäquat durchgeführt?	n.r.
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	n.r.
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	0
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	0
33. Wurden die Diskontraten angegeben?	n.r.
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	0
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	0
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	1
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	1
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	1
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	1
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	0
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	0
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	0
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	1
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	½
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	0
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	½
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	½
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	0
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	½
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	0
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	½
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	0
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	1
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	1

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: AHTAC 1998	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		1
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		1
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		1
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		1
6. Wurde die Zielpopulation klar beschrieben?		1
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		½
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		0
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		0
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		½
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		½
13. Wurden die relevanten Annahmen explizit formuliert?		1
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		½
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		½
16. Wurden adäquate Quellen für die Gesundheitseffektdaten gewählt und eindeutig genannt?		1
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		½
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		n.r.
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		½
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		0
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		0
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		0
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		0
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	½
29. Wurden Währungskonversionen adäquat durchgeführt?	0
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	0
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	0
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	1
33. Wurden die Diskontraten angegeben?	n.r.
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	0
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	0
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	1
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	½
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	1
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	1
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	½
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	0
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	1
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	1
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	½
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	0
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	1
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	½
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	0
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	½
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	½
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	1
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	½
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	1
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	1

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: Radensky et al. 1998	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		½
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		0
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		1
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		0
6. Wurde die Zielpopulation klar beschrieben?		1
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		1
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		0
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		1
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		1
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		1
13. Wurden die relevanten Annahmen explizit formuliert?		1
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		½
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		1
16. Wurden adäquate Quellen für die Gesundheitseffekt-daten gewählt und eindeutig genannt?		1
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		n.r.
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		½
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		1
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		0
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		½
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		½
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		½
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	1
29. Wurden Währungskonversionen adäquat durchgeführt?	n.r.
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	1
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	1
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	1
33. Wurden die Diskontraten angegeben?	1
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	1
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	1
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	½
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	½
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	1
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	1
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	0
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	0
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	0
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	1
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	½
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	0
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	½
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	½
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	½
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	½
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	½
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	0
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	0
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	½
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	½

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: Brown & Garber 1999	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		1
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		1
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		1
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		1
6. Wurde die Zielpopulation klar beschrieben?		1
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		1
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		1
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		1
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		1
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		1
13. Wurden die relevanten Annahmen explizit formuliert?		1
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		½
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		1
16. Wurden adäquate Quellen für die Gesundheitseffekt-daten gewählt und eindeutig genannt?		½
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		n.r.
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		½
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		½
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		½
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		1
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		1
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		1
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	1
29. Wurden Währungskonversionen adäquat durchgeführt?	n.r.
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	1
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	1
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	1
33. Wurden die Diskontraten angegeben?	1
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	1
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	1
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	½
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	1
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	1
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	1
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	1
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	1
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	½
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	½
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	0
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	½
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	1
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	1
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	½
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	0
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	0
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	1
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	1
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	1

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: McCrorry et al. 1999	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		1
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		1
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		1
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		1
6. Wurde die Zielpopulation klar beschrieben?		1
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		1
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		1
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		1
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		1
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		1
13. Wurden die relevanten Annahmen explizit formuliert?		1
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		1
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		1
16. Wurden adäquate Quellen für die Gesundheitseffektdaten gewählt und eindeutig genannt?		1
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		n.r.
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		1
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		1
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		1
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		1
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		1
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		1
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	1
29. Wurden Währungskonversionen adäquat durchgeführt?	n.r.
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	1
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	1
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	1
33. Wurden die Diskontraten angegeben?	1
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	1
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	1
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	1
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	1
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	1
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	1
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	1
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	1
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	1
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	1
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	1
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	1
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	1
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	1
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	1
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	1
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	½
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	1
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	1
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	1

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: Raab et al. 1999	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		1
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		0
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		0
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		0
6. Wurde die Zielpopulation klar beschrieben?		½
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		1
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		1
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		1
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		½
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		1
13. Wurden die relevanten Annahmen explizit formuliert?		1
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		1
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		1
16. Wurden adäquate Quellen für die Gesundheitseffektdaten gewählt und eindeutig genannt?		½
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		n.r.
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		1
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		½
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		½
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		1
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		½
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		½
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	1
29. Wurden Währungskonversionen adäquat durchgeführt?	n.r.
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	0
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	½
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	½
33. Wurden die Diskontraten angegeben?	1
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	0
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	1
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	1
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	1
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	1
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	1
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	0
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	0
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	½
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	1
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	1
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	0
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	½
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	½
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	½
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	½
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	½
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	0
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	0
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	1
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	1

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: Smith et al. 1999	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		1
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		1
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		1
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		1
6. Wurde die Zielpopulation klar beschrieben?		1
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		1
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		1
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		1
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		1
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		1
13. Wurden die relevanten Annahmen explizit formuliert?		1
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		1
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		1
16. Wurden adäquate Quellen für die Gesundheitseffektdaten gewählt und eindeutig genannt?		½
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		n.r.
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		½
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		1
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		0
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		½
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		½
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		½
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	1
29. Wurden Währungskonversionen adäquat durchgeführt?	n.r.
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	½
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	½
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	1
33. Wurden die Diskontraten angegeben?	1
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	0
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	½
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	½
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	1
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	0
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	1
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	1
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	0
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	0
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	1
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	1
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	0
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	½
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	½
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	0
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	1
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	½
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	0
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	0
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	1
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	1

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: Hutchinson et al. 2000	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		1
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		1
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		1
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		1
6. Wurde die Zielpopulation klar beschrieben?		1
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		1
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		1
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		1
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		1
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		1
13. Wurden die relevanten Annahmen explizit formuliert?		½
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		½
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		1
16. Wurden adäquate Quellen für die Gesundheitseffektdaten gewählt und eindeutig genannt?		1
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		n.r.
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		1
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		½
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		½
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		½
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		1
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		½
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	1
29. Wurden Währungskonversionen adäquat durchgeführt?	0
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	0
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	1
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	1
33. Wurden die Diskontraten angegeben?	1
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	1
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	1
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	1
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	1
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	1
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	1
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	0
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	0
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	½
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	½
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	½
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	0
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	½
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	½
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	0
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	½
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	0
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	0
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	½
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	½
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	1

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: Payne et al. 2000	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		1
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		1
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		1
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		1
6. Wurde die Zielpopulation klar beschrieben?		1
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		1
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		1
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		1
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		1
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		½
13. Wurden die relevanten Annahmen explizit formuliert?		1
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		1
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		1
16. Wurden adäquate Quellen für die Gesundheitseffektdata gewählt und eindeutig genannt?		1
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		n.r.
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		1
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		1
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		½
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		½
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		½
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		½
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	1
29. Wurden Währungskonversionen adäquat durchgeführt?	n.r.
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	1
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	½
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	1
33. Wurden die Diskontraten angegeben?	1
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	½
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	1
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	1
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	1
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	1
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	1
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	1
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	0
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	1
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	1
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	0
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	1
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	½
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	0
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	1
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	½
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	½
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	1
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	1
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	1

Checkliste methodische Qualität		
Autoren, Titel und Publikationsorgan: Montz et al. 2001	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung präzise formuliert?		1
2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		1
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben?		1
4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen?		1
5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet?		1
6. Wurde die Zielpopulation klar beschrieben?		1
7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben?		½
8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?		½
9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht?		1
10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		1
Analysemethoden und Modellierung		
11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben?		1
12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)?		½
13. Wurden die relevanten Annahmen explizit formuliert?		½
14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		1
Gesundheitseffekte		
15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?		1
16. Wurden adäquate Quellen für die Gesundheitseffektdaten gewählt und eindeutig genannt?		1
17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)		n.r.
18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)		½
19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?		0
20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?		n.r.
21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)		1
Kosten		
22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?		½
23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?		1
24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?		½
25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?		1
26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?		½

27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?	n.r.
28. Wurde die Währung genannt?	1
29. Wurden Währungskonversionen adäquat durchgeführt?	n.r.
30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?	0
Diskontierung	
31. Wurden zukünftige Gesundheitseffekte und Kosten adäquat diskontiert?	1
32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?	1
33. Wurden die Diskontraten angegeben?	1
34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?	1
Ergebnispräsentation	
35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?	1
36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?	½
37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?	½
38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?	1
39. Wurden reine (nicht lebensqualitätskorrigierte) klinische Effekte berichtet?	1
40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?	1
41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?	0
Behandlung von Unsicherheiten	
42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	1
43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?	0
44. Wurde Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?	½
45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?	½
46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?	½
47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?	0
Diskussion	
48. Wurde die Datenqualität kritisch beurteilt?	½
49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?	½
50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?	0
51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?	½
52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?	1
53. Wurden wichtige ethische und Verteilungsfragen diskutiert?	½
54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?	½
Schlussfolgerungen	
55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?	1
56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?	1

II 3 Studienmerkmale und Ergebnisse der eingeschlossenen gesundheitsökonomischen Studien

Tabelle 47: Studienmerkmale und Ergebnisse der eingeschlossenen gesundheitsökonomischen Studien

Quelle	Publikationstyp	Studientyp	Land/Währung	Studienpopulation	Fragestellung	Technik	Ergebnisse	Übertragbarkeit	Kommentar / Limitierung
Schechter, 1996	Zeitschriftenartikel	Kosten-Effektivitäts-Analyse	U.S.A./US\$ (1994)	Hypothetische Kohorte von 20-64-jährigen Frauen, die nach dem US-Zensus von 1990 verteilt sind	Wie kosteneffektiv ist ein Rescreening von konventionell aufbereiteten Zervixabstrichen mit PAPNET®?	Adaptiertes Markov-Modell (nach Müller et al., 1990)	Bei 30% höherer Sensitivität ergibt sich für PAPNET® eine C/E1-Ratio von US\$48,474 für ein 2-Jahres-Screeningintervall und von US\$25,185 für ein 3-Jahres-Screeningintervall. Folglich ist PAPNET®-Rescreening ab 2-Jahresintervall kosteneffektiver als konventionelle Methode	Nein (da sehr US-spezifisch (Altersverteilung der Bevölkerung, Kosten, Screeningintervalle zudem Differenz zum deutschen Gesundheitssystem)	Umfassende C/E-Studie Limitierung: Sensitivitäts-, Spezifitätswerte für PAPNET® (30% mehr Sensitivität als für konventionelle Methode, Spezifität=95%) stammen aus Herstellerangaben; Autor ist „consultant“ von Neuromedical Systems Inc., dem Hersteller von PAPNET®
Noorani et al., 1997	HTA-Bericht	Unsystematischer Review, Kosten-Effektivitäts-Analyse	Kanada/\$ (fraglich ob US\$ oder CAN\$?)	Hypothetische „Kohorte“ von 4 Millionen Pap-Abstrichen, entsprechend der jährlich in Kanada gescreent Population	Kosten-Effektivität im Vergleich von 3 Rescreening-Methoden: (1) AutoPap® 300QC, (2) PAPNET®, (3) 10%iges Zufallscreening mit konventioneller Methode	Entscheidungsanalyse (Markov-Modell)	Pro zusätzlich gefundenen Krebsfall: (1) \$250 (2) >\$400 (3) >\$10,000 (2) und (3) sind etwa gleich effektiv (Rescreening Sensitivität (2) 83% vs. (3) 80%), aber deutlich unterschiedlich hinsichtlich der Kosten	Nein (da Datenherkunft unklar, mäßige Studienqualität, anderes Gesundheitssystem)	Kosten/Effektivitätsstudie mit Limitierungen, z. B. Unklarheiten bezüglich der Währung, keine Diskontierung der Kosten und Effekte

Quelle	Publikationstyp	Studientyp	Land/Währung	Studienpopulation	Fragestellung	Technik	Ergebnisse	Übertragbarkeit	Kommentar / Limitierung
AHTAC, 1998	HTA-Bericht	Systematischer Review, Kosten-Effektivitäts-Analyse	Australien/Währung unklar (US\$?)	5.022.752 australische Frauen mit normalem Krebsrisiko zwischen 20 und 74 Jahren (nach Australischem Zensus 1996)	Übersicht zu zusätzlichen Kosten und Nutzen durch den Einsatz der neuen automatisierten und halbautomatisierten Zervixscreeningverfahren	Flow Chart-Modell nach Bragett et al, 1993	Baseline (konventionelles Pap-Screening): \$47.000 / potentiell entdecktem Krebsfall (inklusive Diagnostik und Therapie) pro Zyklus (2 Jahre) Neue Methoden (automatisiert und halbautomatisiert): \$240.000/potentiell neuentdecktem Krebsfall/2-Jahres-Zyklus	Nein (Australische Daten, spezielle auf Australien adaptierte Annahmen)	Neue Methode zu kostenintensiv, um routinemäßig in das Australische Screeningprogramm aufgenommen zu werden, das laut der Autoren bereits auf hohem Niveau läuft - Limitiert durch abweichenden Modellcharakter
Radensky et al., 1998	Zeitschriftenartikel	Kostenanalyse	U.S.A./US\$ (1997)	Hypothetische Kohorte von 20-75-jährigen Frauen (keine Angaben über die Anzahl), die in 3-jährigem Intervall gescreent werden	Welchen ökonomischen Einfluss hat der Einsatz der interactive neural network-assisted (INNA) Screeningverfahren im Zervixscreening?	9-Zustände-Markov-Modell, modifiziert nach Eddy et al., 1990	Bei Sensitivität der konventionellen Pap2-Methode von 85%: Kosten/geretem Lebensjahr zwischen \$35.000 und \$80.000 Analyse sensitiv gegenüber Sensitivität der konventionellen Pap-Methode, aber nicht der Krebsbehandlungskosten	Nein (amerikanisches Gesundheitssystem, Studienqualität)	Keine vollständige Kosten/Effektivitäts-Analyse Fehlen der marginalen Kosten
Raab et al., 1999	Zeitschriftenartikel	Kosten-Effektivitäts-Analyse	U.S.A./US\$	„Referenzfall“ einer 30-jährigen, weißen Frau, 26.352 Abstriche der Universität von Iowa (1995)	Wie ändert sich die Kosten / Effektivität des Zervixscreenings durch die Einführung der neuen Technologien?	Entscheidungsanalytische Modellierung (Entscheidungsbaum)	Damit die neuen Technologien kosteneffektiv werden, müssen entsprechend viele Läsionen (HSL) entdeckt werden, was nicht nachzuweisen war (Literaturangaben), konventionelle Methode selbst mit Behandlung noch kosteneffektiv (US\$19.655 pro zusätzlichem, diskontiertem Lebensjahr)	Nein (US-amerikanische Population, mit entsprechendem zugrundeliegenden Daten und Gesundheitssystem)	Limitiert durch fehlende Konkretisierung der „neuen Verfahren“, Threshold-Analyse der neuen Technologien

Quelle	Publikationstyp	Studientyp	Land/Währung	Studienpopulation	Fragestellung	Technik	Ergebnisse	Übertragbarkeit	Kommentar / Limitierung
Brown & Garber; 1999	Zeitschriftenartikel	Kosten-Effektivitäts-Analyse	U.S.A./US\$ (1996)	Hypothetische Kohorte von 20-65-jährigen Frauen, die repräsentativ für US-Bevölkerung ist	Wie kosteneffektiv sind die drei neuen Technologien (ThinPrep®, PAPNET®, AutoPap®) im Vergleich zum konventionellen Pap-Test?	Entscheidungsanalyse, basierend auf 9-Zustände-Markov-Modell von Eddy, 1990	Alle 3 Technologien verbessern Outcomes (Morbidität, Mortalität, zusätzlich gewonnene Lebensjahre), aber allesamt zu höheren Kosten Lebenserwartung wird z.B. zwischen 5h und 1,6d verbessert (konventioneller Pap-Test immer beste C/E-Ratio) AutoPap® generell bessere Outcomes zu niedrigeren Kosten im Vergleich zu ThinPrep® und PAPNET® (bleibt ab 2-Jahres-Intervall aufwärts als einzige neue Technologie unter US\$5000-Grenze für inkrementelle C/E-Ratio) ThinPrep® wird von allen Technologien in jedem Screeningintervall dominiert C/E-Ratios der neuen Technologien verbessert durch: 1. höhere Prävalenz 2. niedrigere Sensitivität des Pap-Tests 3. höhere Sensitivität der neuen Tests	Fraglich (da US-Bevölkerungs-Standard und Gesundheitssystem)	Umfassende, gut nachvollziehbare C/E-Studie Große Diskrepanz hinsichtlich der Ergebnisse zur Studie von Hutchinson et al., 2000 (dort ThinPrep® beste C/E-Ratio ab 3-Jahres-Intervall)?

Quelle	Publikationstyp	Studientyp	Land/Währung	Studienpopulation	Fragestellung	Technik	Ergebnisse	Übertragbarkeit	Kommentar / Limitierung
Smith et al., 1999	Zeitschriftenartikel	Kosten-Effektivitäts-Analyse	U.S.A./US\$ (1997)	Hypothetische Kohorte von 100.000 amerikanischen 18-jährigen Frauen	Kosten/Effektivität des Einsatzes von AutoPap® im Vergleich mit konventioneller Methode des Zervixscreenings	Entscheidungsanalyse (7-Zustände-Markov-Modell)	AutoPap® immer die bessere C/E-Ratio als konventionelle Methode (z.B. bei jährlichem Screening werden mit AutoPap® 32,1 Tag hinzugewonnen und US\$628 gespart) Je größer das Screeningintervall, um so geringer die Effekte (gerettete Lebensstage) und die gesparten Kosten	Nein (da sehr spezielle Annahmen, kein Referenzstandard, amerikanisches Gesundheitssystem)	C/E-Analyse für spezielle Kohorte, die vielen, teilweise realitätsfernen Annahmen unterliegt Einzige Studie, die geringere C/E-Ratio für neue Technologie analysiert Studie wurde von Neopath Inc., dem Hersteller von AutoPap®, finanziert
McCrony et al., 1999	HTA-Bericht	Systematischer Review, Metaanalyse, Kosten-Effektivitäts-Analyse	U.S.A./US\$	Hypothetische Kohorte von 15-85-jährigen Frauen	Wie ist der Effekt des Einsatzes von liquid-based Technologie und automatisierten Rescreening-Techniken im regulären Zervixscreeningprogramm im Vergleich zur konventionellen Pap-Methode auf die gesamten Gesundheitskosten, Morbidität und Mortalität	20-Zustände-Markov-Modell	Inkrementelle Kosten/gerettetem Lebensjahr in 3-Jahres-Intervall: → (1) liquid based Technologie versus Pap-Methode: US\$22.010 → (2) 100% Rescreening versus Pap-Methode: US\$24.988 Inkrementelle Kosten/gerettetem Lebensjahr in 2-Jahres-Intervall: → (1) US\$45.265 → (2) US\$17.3484 Inkrementelle Kosten/gerettetem Lebensjahr jährliches Screening: → (1) US\$173484 → (2) US\$227343	Mit Einschränkung (da auf US-amerikanisches Gesundheitssystem ausgerichtet)	Ausführlichstes Modell mit umfassender Kosten-Effektivitäts-Analyse, Kontextdokument=[255]

Quelle	Publikationstyp	Studientyp	Land/Währung	Studienpopulation	Fragestellung	Technik	Ergebnisse	Übertragbarkeit	Kommentar / Limitierung
Payne et al., 2000	HTA-Bericht	Systematischer Review	Großbritannien/£ (1999)	100.000 Frauen im Alter von 18 bis 95 Jahren	Auswirkung der neuen liquid-based Technologie hinsichtlich Zervixkrebsinzidenz, -mortalität, -kosten Vergleich der Kosten-Effektivität der liquid-based Technologie mit konventioneller Pap-Methode)	Entscheidungsanalysemodell	Je geringer Screeningabstand, um so kleiner Inzidenz, Morbidität und Mortalität bei beiden Technologien Inkrementelle gewonnene Lebensstage bei 3-Jahres-Screening (1): 0,26, bei 2-Jahres-Screening (2): 0,33 Inkrementelle Kosten/gerettetem Lebensjahr bei (1) : £2,30, bei (2): £3,15 Inkrementelle Kosten/gerettetem Lebensjahr bei (1) : £2,544, bei (2): £4,422	Modell im Ganzen nicht (da sehr spezifisch auf GB ausgerichtet (Diskontrate, Beteiligungsrate von 85%, etc.) darunter sind mit beiden Technologien nicht kosteneffektiv insgesamt sehr umfassende C/E-Studie, aber sehr GB-spezifisch und mit sehr sensiblen Parametern (Siehe Sensitivitätsanalysen)	Je mehr Nutzen, um so teurer Screeninguntersuchungen im Abstand von 2 Jahren oder darunter sind mit beiden Technologien nicht kosteneffektiv insgesamt sehr umfassende C/E-Studie, aber sehr GB-spezifisch und mit sehr sensiblen Parametern (Siehe Sensitivitätsanalysen)
Hutchinson et al., 2000	Zeitschriftenartikel	Kohortenstudie, Kosten-Effektivitäts-Analyse	U.S.A./US\$ (1997)	Hypothetische Kohorte von 100.000 Frauen im Alter von 20 bis 65 Jahren	Vergleich der verfügbaren Zervixscreening-Techniken, inklusive verschiedener neuer Verfahren, unter aktuellen Bedingungen	Markov-Modellierung	Neue Technologien können die Anzahl der Krebsfälle um 45% bis 60% reduzieren (bei 100%iger Compliance) Bei 2-jährigem Intervall oder größer ergibt sich Kosten/Effektivitäts-Verhältnis unter 50.000 \$ pro gerettetem Lebensjahr (für ThinPrep®) Neue Techniken (ThinPrep®) können, aufgrund der verbesserten Sensitivität, die Krebsinzidenz auf 9,5/100.000 gescreenten Frauen senken (im Vergleich zu 13,2/100.000 Frauen mit konventioneller Methode)	Mit Einschränkungen (da auf US amerikanisches Gesundheitssystem ausgerichtet)	Umfassende und qualitativ hochwertige Kosten/Effektivitäts-Analyse Limitierungen: nur direkte Kosten werden analysiert, Lebensqualität wird nicht betrachtet, Studie wurde von CYTYC Corp. (Hersteller von ThinPrep®) finanziert

Quelle	Publikationstyp	Studientyp	Land/Währung	Studienpopulation	Fragestellung	Technik	Ergebnisse	Übertragbarkeit	Kommentar / Limitierung
Montz et al. 2001	Zeitschriftenartikel	Kosten-Effektivitäts-Analyse	U.S.A./US\$ (1997)	Hypothetische Kohorte von 100.000 Frauen im Alter von 20 bis 80 Jahren	Vergleich liquid-based und konventioneller Pap-Test unter 3 verschiedenen Compliance Szenarien: 1. selbst angegebene Compliance, 2. Healthy people 2000, 3. Healthy people 2010	Markov-Modellierung (Modell Hutchinsonson 2000)	Zervixkarzinominzidenz wird reduziert um 32% durch liquid-based Screeningverfahren bei derzeitiger Compliance, um 50% unter Compliance Nr. 3, Pap-Test reduziert Inzidenz um 10,9% (Nr. 2) und 22,9% (Nr. 3). Liquid-based Verfahren im Vergleich zum Pap-Test: diskontierte inkrementelle KER von 15.296 US\$/LYS (Nr. 1), 16.796 US\$/LYS (Nr. 2), 20.424 US\$/LYS (Nr. 3)	Mit Einschränkungen (da auf US-amerikanisches Gesundheitssystem ausgerichtet)	Finanzierung durch Cytoc

II 4 Einzelergebnisse der inhaltlich-qualitativen Bewertung evaluierter ökonomischer Studien

Tabelle 48: Inhaltlich-qualitative Bewertung der evaluierten ökonomischer Studien, Einzelergebnisse.

Qualitätskriterien	Schechter 1996	Noorani et al. 1997	AHTAC 1998	Radensky et al. 1998	Brown & Garber 1999	McCroory et al. 1999	Raab et al. 1999	Smith et al. 1999	Hutchinson et al. 2000	Payne et al. 2000	Montz et al. 2001	Mittelwert (% der max. Punktzahl)
Fragestellung												
1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	100
2	1.0	0.5	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	91
Summe	2.0	1.5	2.0	1.5	2.0	2.0	2.0	2.0	2.0	2.0	2.0	
Evaluationsrahmen												
3	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	82
4	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	91
5	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	82
6	1.0	0.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	1.0	1.0	86
7	1.0	0.5	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	86
8	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	77
9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	100
10	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	91
Summe	8.0	6.5	5.5	5.0	8.0	8.0	4.5	8.0	8.0	8.0	7.0	

Qualitätskriterien	Schechter 1996	Noorani et al. 1997	AHTAC 1998	Radensky et al. 1998	Brown & Garber 1999	McCroy et al.	Raab et al. 1999	Smith et al. 1999	Hutchinson et al. 2000	Payne et al. 2000	Montz et al. 2001	Mittelwert (% der max. Punktzahl)
Analysemethoden und Modellierung												
11	1.0	0.5	0.5	1.0	1.0	1.0	0.5	1.0	1.0	1.0	1.0	86
12	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5	86
13	0.5	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.5	1.0	0.5	82
14	1.0	0.5	0.5	0.5	0.5	1.0	1.0	1.0	0.5	1.0	1.0	77
Summe	3.5	2.5	2.5	3.5	3.5	4.0	3.5	4.0	3.0	3.5	3.0	
Gesundheitseffekte												
15	1.0	0.5	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	91
16	0.5	0.5	1.0	1.0	0.5	1.0	0.5	0.5	1.0	1.0	1.0	77
17	0.5	n.r.	0.5	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	50
18	n.r.	1.0	n.r.	0.5	0.5	1.0	1.0	0.5	1.0	1.0	0.5	88
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
20	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.
21	1.0	0.5	0.5	1.0	0.5	1.0	0.5	1.0	0.5	1.0	1.0	77
Summe	3.0	2.5	2.5	3.5	2.5	4.0	3.0	3.0	3.5	4.0	3.5	
Kosten												
22	0.5	0.5	0.0	0.0	0.5	1.0	0.5	0.0	0.5	0.5	0.5	41
23	1.0	0.5	0.0	0.5	1.0	1.0	1.0	0.5	0.5	0.5	1.0	68
24	0.5	0.5	0.0	0.5	1.0	1.0	0.5	0.5	1.0	0.5	0.5	59
25	0.5	0.5	0.0	0.5	1.0	1.0	0.5	0.5	0.5	0.5	1.0	59
26	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	50
27	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.
28	1.0	0.5	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	91
29	n.r.	n.r.	0.0	n.r.	n.r.	n.r.	n.r.	n.r.	0.0	n.r.	n.r.	0
30	1.0	n.r.	0.0	1.0	1.0	1.0	0.0	0.5	0.0	1.0	0.0	50
Summe	5.0	3.0	1.0	4.0	6.0	6.5	4.0	3.5	4.0	4.5	4.5	

Qualitätskriterien	Schechter 1996	Noorani et al. 1997	AHTAC 1998	Radensky et al. 1998	Brown & Garber 1999	McCroy et al. 1999	Raab et al. 1999	Smith et al. 1999	Hutchinson et al. 2000	Payne et al. 2000	Montz et al. 2001	Mittelwert (% der max. Punktzahl)
Diskontierung												
31	1.0	0.0	0.0	1.0	1.0	1.0	0.5	0.5	1.0	0.5	1.0	68
32	1.0	0.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	1.0	1.0	86
33	1.0	n.r.	n.r.	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	90
34	0.5	0.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	0.5	1.0	55
Summe	3.5	0.0	1.0	4.0	4.0	4.0	2.0	2.5	4.0	3.0	4.0	
Ergebnispräsentation												
35	1.0	0.0	0.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	1.0	77
36	0.0	1.0	1.0	0.5	0.5	1.0	1.0	0.5	1.0	1.0	0.5	73
37	1.0	1.0	0.5	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.5	86
38	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	100
39	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	91
40	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	91
41	1.0	0.0	0.5	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	50
Summe	5.0	5.0	5.0	5.0	6.5	7.0	6.0	5.0	6.0	7.0	5.0	
Behandlung von Unsicherheiten												
42	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	100
43	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	27
44	1.0	0.0	1.0	0.0	1.0	1.0	0.5	0.0	0.5	0.0	0.5	50
45	1.0	1.0	1.0	1.0	0.5	1.0	1.0	1.0	0.5	1.0	0.5	86
46	0.5	0.5	0.5	0.5	0.5	1.0	1.0	1.0	0.5	1.0	0.5	68
47	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	18
Summe	4.5	2.5	3.5	2.5	4.0	6.0	3.5	3.0	2.5	4.0	2.5	

Qualitätskriterien	Schechter 1996	Noorani et al. 1997	AHTAC 1998	Radensky et al. 1998	Brown & Garber 1999	McCroy et al. 1999	Raab et al. 1999	Smith et al. 1999	Hutchinson et al. 2000	Payne et al. 2000	Montz et al. 2001	Mittelwert (% der max. Punktzahl)
Diskussion												
48	0.5	0.5	1.0	0.5	0.5	1.0	0.5	0.5	0.5	1.0	0.5	64
49	0.5	0.5	0.5	0.5	1.0	1.0	0.5	0.5	0.5	0.5	0.5	59
50	0.0	0.0	0.0	0.5	1.0	1.0	0.5	0.0	0.0	0.0	0.0	27
51	0.5	0.5	0.5	0.5	0.5	1.0	0.5	1.0	0.5	1.0	0.5	64
52	0.0	0.0	0.5	0.5	0.0	1.0	0.5	0.5	0.0	0.5	1.0	41
53	0.5	0.5	1.0	0.0	0.0	0.5	0.0	0.0	0.0	0.5	0.5	32
54	1.0	0.0	0.5	0.0	1.0	1.0	0.0	0.0	0.5	1.0	0.5	50
Summe	3.0	2.0	4.0	2.5	4.0	6.5	2.5	2.5	2.0	4.5	3.5	
Schlussfolgerungen												
55	0.5	1.0	1.0	0.5	1.0	1.0	1.0	1.0	0.5	1.0	1.0	86
56	1.0	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	95
Summe	1.5	2.0	2.0	1.0	2.0	2.0	2.0	2.0	1.5	2.0	2.0	
Summe Gesamt	39.0	27.5	29.0	32.5	42.5	50.0	33.0	35.5	36.5	42.5	37.0	

III Systematische Entscheidungsanalyse

Verfasser:

Uwe Siebert^{1,2,3}, Gaby Sroczynski^{1,2}, Bernhard Gibis⁴, Pamela Aidelsburger⁵, Jutta Engel⁶, Peter Hillemanns⁷, Kerstin Voigt¹, Christiane Muth⁸, Marcial Velasco-Garrido⁹, Jürgen Wasem⁵, Dieter Hölzel⁶, Sue J. Goldie²

- ¹ Bayerische Forschungs- und Koordinierungsstelle Public Health, Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Klinikum Großhadern, Ludwig-Maximilians-Universität, München
- ² Harvard Center for Risk Analysis, Harvard School of Public Health, Boston, USA
- ³ Institute for Technology Assessment, Massachusetts General Hospital, Harvard Medical School, Boston, USA
- ⁴ Kassenärztliche Bundesvereinigung (KBV), Dezernat 2, Versorgungsqualität und Sicherstellung, Berlin
- ⁵ Alfried Krupp von Bohlen und Halbach-Stiftungslehrstuhl für Medizinmanagement, Universität Duisburg-Essen, Essen
- ⁶ Tumorzentrum München, Tumorregister, Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Klinikum Großhadern, Ludwig-Maximilians-Universität, München
- ⁷ Klinik und Poliklinik für Frauenheilkunde und Geburtshilfe, Klinikum Großhadern, München
- ⁸ Abteilung Epidemiologie, Sozialmedizin und Gesundheitssystemforschung, Medizinische Hochschule Hannover, Hannover
- ⁹ Institut für Gesundheitswissenschaften, FG Management im Gesundheitswesen, Technische Universität Berlin

III C Hauptdokument Systematische Entscheidungsanalyse

III C 1 Policy Question

Die Policy Question für den gesamten HTA-Bericht wurde in Kapitel I C 1 beschrieben.

III C 2 Hintergrund

Der Hintergrund für den gesamten HTA-Bericht mit epidemiologischen, medizinischen, ökonomischen und entscheidungsbezogenen Aspekten wurde detailliert in Kapitel I C 2 dargestellt.

III C 3 Forschungsfragen

Ziel der systematischen Entscheidungsanalyse ist es, die medizinischen und ökonomischen Langzeit-Konsequenzen des Einsatzes neuer Screeningtechnologien (Dünnschichtpräparationen und automatisierte Auswertungsverfahren) in der Zervixkrebsfrüherkennung unter Einbeziehung der in den vorangegangenen Kapiteln gesammelten Evidenz unter der für Deutschland spezifischen klinischen und gesetzlich festgelegten Praxis zu modellieren und die Relevanz der derzeitigen Informationsunsicherheit zu beurteilen.

Die zu beantwortenden Forschungsfragen lauten:

1. Welche Auswirkung hat der Einsatz unterschiedlicher Strategien von Primärscreening-Technologien im Kontext des deutschen Krebsfrüherkennungsprogramms auf die Kosten und Lebenserwartung?
2. Wie ist das Kosten-Effektivitäts-Verhältnis der untersuchten Technologien im Vergleich zum Standardverfahren (Papanicolaou-Verfahren mit manueller zytologischer Auswertung)?
3. Wie würde sich die Kosten-Effektivität bei verschiedenen Screeningintervallen verhalten?
4. In welchen Bereichen besteht weiterer Forschungsbedarf?

III C 4 Methodik

Zur Evaluation der Auswirkungen des Technologieeinsatzes auf die Lebenserwartung und Langzeit-Kosten wurde eine medizinische Entscheidungsanalyse und eine Kosten-Effektivitäts-Analyse durchgeführt. Kosten-Effektivitäts-Analysen sind die Methode der Wahl, um die medizinischen und ökonomischen Langzeit-Konsequenzen eines Technologieeinsatzes in quantitative Beziehung zu setzen (Gold et al. 1996; Leidl et al. 1999). Ziel der Methode ist es, Entscheidungsträger bei der rationalen Allokation von Ressourcen zu unterstützen (Wasem & Siebert 1999).

III C 4. 1 Methodik der Markov-Modellierung

Im Rahmen dieses HTA-Projekts wurde ein Markov-Modell (*state transition model*) entwickelt, das den natürlichen Krankheitsverlauf und den Krankheitsverlauf unter den verschiedenen Screeningstrategien mit den jeweils anfallenden Kosten für Screening, Diagnostik und Behandlung einer Kohorte von Frauen im Alter von 15 Jahren über deren Lebenszeit abbildet.

Ein Markov-Modell enthält verschiedene a priori festgesetzte Gesundheitszustände (Markov-Zustände), die von einer hypothetischen Kohorte von Patienten durchlaufen werden können. Dabei wird die Zeit in feste Zeitintervalle (Markov-Zyklen), z.B. Jahresintervalle, eingeteilt. Nach jedem Zyklus können die Mitglieder der Kohorte mit bestimmten Übergangswahrscheinlichkeiten entweder in andere Gesundheitszustände übertreten, im aktuellen Gesundheitszustand verweilen oder an der untersuchten Krankheit oder anderen Ursachen sterben. In einem Markov-Modell hängen die Übergangswahrscheinlichkeiten ausschließlich vom momentanen Gesundheitszustand ab (Markov-Annahme) (Beck & Pauker 1983; Sonnenberg & Beck 1993; Siebert U. & Kurth 2000).

In der sogenannten Markov-Spur werden für die Kohorte pro Zyklus die Effektivität (klinische Ereignisse, Lebensjahre) und die angefallenen Kosten festgehalten. Die Erwartungswerte für Kosten und medizinische Effektivität ergeben sich durch Aufsummierung der zyklenspezifischen Werte über den gesamten Zeithorizont der Analyse und können für die verschiedenen evaluierten medizinischen Technologien gegenübergestellt werden. Aus dem Quotienten der Kostendifferenz (inkrementelle Kosten) und der Effektivitätsdifferenz (inkrementelle Effektivität) ergibt sich dann das

inkrementelle Kosten-Effektivitäts-Verhältnis (IKEV) einer Technologie im Vergleich zu einer anderen (Leidl 1998). Aber auch für die rein klinische Entscheidungsanalyse zur Abwägung von Lebensdauer und Lebensqualität eignet sich der Einsatz von Markov-Modellen (Siebert U. & Kurth 2000).

Markov-Modelle sind als entscheidungsanalytische Methode anzuwenden, wenn mindestens eines der folgenden Kriterien erfüllt ist (Sloan 1995):

1. Das Entscheidungsproblem beinhaltet zeitveränderliche Risiken oder Kosten.
2. Der Zeitpunkt des Eintretens eines bestimmten Ereignisses (time-to-event) spielt eine Rolle.
3. Relevante Ereignisse können mehrmals auftreten.

Alle drei Kriterien sind im Falle der Entstehung und des Verlaufs des Zervixkarzinoms erfüllt und erfordern damit die Verwendung eines Markov-Modells zur Analyse der vorliegenden Forschungsfragen.

Das im Rahmen dieses HTA verwendete Markov-Modell wird weiter unten im Detail beschrieben (III C 4. 8)

III C 4. 2 Evaluierete Strategien

In der Entscheidungsanalyse im Rahmen dieses HTA wurde bewusst auf eine getrennte Analyse von Produkten unterschiedlicher Hersteller der Dünnschichtpräparationsverfahren bzw. der computergestützten Auswertungsverfahren verzichtet. Um ein für Deutschland möglichst umfassendes Bild zu erhalten, wurden im Rahmen der medizinischen und gesundheitsökonomischen Entscheidungsanalyse die folgenden Strategien evaluiert:

1. Kein Screening
2. Screening mit Papanicolaou-Verfahren plus manuelle zytologische Auswertung
3. Screening mit Dünnschichtpräparation plus manuelle zytologische Auswertung

4. Screening mit Papanicolaou-Verfahren plus automatisierte (computergestützte) zytologische Auswertung
5. Screening mit Dünnschichtpräparation plus automatisierte (computergestützte) zytologische Auswertung

In der Basisfall-Analyse wurde entsprechend der Praxis nach dem gesetzlich festgelegten Krebsfrüherkennungsprogramms in Deutschland (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) 2000) ein einjähriges Screeningintervall verwendet. In weiteren Szenarien wurden jedoch zusätzlich Analysen mit Screeningintervallen von zwei, drei und fünf Jahren durchgeführt. Die Basisfall-Analyse bezog sich auf Frauen, die regelmäßig am einjährigen Screening teilnehmen. Auf Populationsebene bedeutet dies die Annahme einer 100%igen Teilnehmerate der Frauen am Screeningprogramm. Diese Annahme wurde in Sensitivitätsanalysen variiert.

Eine Screeninguntersuchung wurde im Modell gemäß des gesetzlich festgelegten Krebsfrüherkennungsprogramms in Deutschland erstmalig im Alter von 20 Jahren durchgeführt. Entsprechend der Regelungen für das gesetzlich festgelegte Krebsfrüherkennungsprogramm in Deutschland wurde keine Altersbeschränkung für das Screening angenommen. Das Start-Alter für die Modellkohorte wurde auf 15 Jahre festgesetzt. Damit wird implizit angenommen, dass es vor einem Alter von 15 Jahren keine relevante Inzidenz von progredienten intraepitheliale Läsionen oder Zervixkarzinomen gibt und eine solche Progression erst ab dem 15. Lebensjahr stattfindet. Progressionen zwischen dem 15. Lebensjahr (Beginn des analysierten Zeithorizonts) und dem 20. Lebensjahr (Beginn des Screenings) können zu prävalenten Läsionen und Zervixkarzinomfällen führen, die durch das Screeningverfahren ab dem 20. Lebensjahr entdeckt werden können.

III C 4. 3 Zielgrößen der Entscheidungsanalyse

Medizinische Zielgrößen (Outcomes) der Analyse waren:

- Restlebenserwartung (in Lebensjahren, LJ)
- Entdeckte Krebsfälle
- Vermiedene Krebsfälle
- Absolute Anzahl Krebstodesfälle

- Vermiedene Krebstodesfälle

Ökonomische Zielgrößen waren:

- Lebenslange direkte Kosten (in €)
- Kosten-Effektivitäts-Verhältnis (in €/gewonnene Lebensjahre)

Vor der Berechnung der inkrementellen medizinischen Effekte und Kosten für jede Strategie wurden Strategien aufgrund von Dominanz und erweiterter Dominanz eliminiert. Unter Dominanz versteht man den Fall, dass eine Strategie weniger effektiv und gleichzeitig teurer ist als eine andere Therapie. Unter erweiterter Dominanz versteht man den Fall, dass eine Strategie weniger effektiv und teurer ist als eine gewichtete Linearkombination zweier anderer Strategien.

Nach Elimination der dominierten und erweitert dominierten Strategien wurde das inkrementelle Kosten-Effektivitäts-Verhältnis jeder Strategie ermittelt als inkrementelle diskontierte Kosten (in €) dividiert durch inkrementelle diskontierte Restlebenserwartung (in gewonnenen Lebensjahren, LJ). Die Inkremente sind dabei immer als Differenz zwischen dem Ergebnis einer bestimmten Strategie und dem Ergebnis der bezüglich der Effektivität nächst weniger effektiven Strategie zu verstehen. Das inkrementelle Kosten-Effektivitäts-Verhältnis in € pro gewonnene Lebensjahre (€/LJ) ist zu interpretieren als die zusätzlichen Kosten, die für eine Erhöhung der Lebenserwartung um ein zusätzliches Lebensjahr aufgebracht werden müssen.

III C 4. 4 Zeithorizont

Es wurde beginnend mit einem Alter von 15 Jahren ein lebenslanger Zeithorizont analysiert. Diese Angabe betrifft den Analysezeitraum und nicht den Screeningzeitraum. Zu letzterem siehe den Abschnitt III C 4. 2 Evaluierete Strategien.

III C 4. 5 Perspektive

Diese Analyse nimmt die gesamtgesellschaftliche Perspektive ein.

III C 4. 6 Diskontierung

Kosten und Lebenserwartung wurden mit einer jährlichen Rate von 3% diskontiert (Gold et al. 1996). Die Diskontrate wurde in einer Sensitivitätsanalyse zwischen 0%-10% variiert (Anonymous 1999a).

III C 4. 7 Gesellschaftliche Zahlungsbereitschaft

Für Deutschland ist nach Kenntnis der Autoren kein Schwellenwert für die Kosten-Effektivität medizinischer Technologien im Sinne einer gesellschaftlichen Zahlungsbereitschaft publiziert (Wasem 1999; Wasem & Siebert 1999). Dies wurde in einem Survey bestätigt, bei dem 30 deutsche Experten aus dem Bereich der Gesundheitsökonomie befragt wurden (Survey im Jahr 2001, durchgeführt von einem der Autoren, U.S.). Aus diesem Grunde ist die Kosten-Effektivität einer Technologie anhand der berichteten Kosten-Effektivitäts-Verhältnisse von jedem Entscheidungsträger selbst im jeweiligen Kontext zu beurteilen (Leidl 1998).

Im Rahmen des vorliegenden HTA-Reports wurde im Sinne einer Arbeitsdefinition eine Technologie als kosteneffektiv bezeichnet, wenn das Kosten-Effektivitäts-Verhältnis unter 50.000 €/LJ lag. Diese Definition erfolgte in Anlehnung an US-Studien zu Technologien, die auch in Deutschland einen gesellschaftlich akzeptierten Stellenwert besitzen, wie beispielsweise die Hämodialyse (Weinstein 1999) oder die koronare Bypass-Operation (Wong et al. 1990).

In allen Fällen wurden in diesem Bericht jedoch die absoluten Werte der Kosten-Effektivitäts-Verhältnisse dargestellt, um dem Entscheidungsträger eine Bewertung unabhängig von bestimmten Schwellenwerten zu ermöglichen.

III C 4. 8 Beschreibung der Modellstruktur

Es wurde ein entscheidungsanalytisches Markov-Modell entwickelt, in dem eine hypothetische Kohorte von Frauen mit einem initialen Alter von 15 Jahren verschiedene Gesundheitszustände durchläuft, die basierend auf den zytologisch-histologischen Stadien der Zervixerkkrankung festgelegt wurden (Abbildung 11).

Folgende Gesundheitszustände wurden definiert:

1. Keine Zervixzellläsion (Gesund)
2. Zustand nach benigner Hysterektomie
3. Milde zervikale intraepitheliale Neoplasien (CIN 1)
4. Moderate zervikale intraepitheliale Neoplasien (CIN 2)
5. Schwere zervikale intraepitheliale Neoplasien (CIN 3) / Carcinoma in situ
6. Invasives Zervixkarzinom FIGO-Stadium I
7. Invasives Zervixkarzinom FIGO-Stadium II
8. Invasives Zervixkarzinom FIGO-Stadium III
9. Invasives Zervixkarzinom FIGO-Stadium IV
10. Zustand Lebend nach 5 Jahren post Diagnose eines Zervixkarzinom FIGO I (Survivor FIGO I)
11. Zustand Lebend nach 5 Jahren post Diagnose eines Zervixkarzinom FIGO II (Survivor FIGO II)
12. Zustand Lebend nach 5 Jahren post Diagnose eines Zervixkarzinom FIGO III (Survivor FIGO III)
13. Zustand Lebend nach 5 Jahren post Diagnose eines Zervixkarzinom FIGO IV (Survivor FIGO IV)
14. Tod durch Zervixkrebs
15. Tod aus anderer Ursache

Für die Zustände 6-9 (FIGO-Stadium I-IV) wurden im Markov-Modell je ein Zustand für den diagnostischen Status "Karzinom entdeckt" und "Karzinom nicht entdeckt" verwendet, so dass das Markov-Modell insgesamt 19 Zustände umfasst.

Der lebenslange Zeithorizont der Analyse wurde in gleichgroße zeitliche Intervalle (Markov-Zyklus) unterteilt, in denen die Frauen von einem zum anderen Gesundheitszustand übertreten bzw. im aktuellen Zustand verweilen konnten. Als Zykluslänge wurde ein Zeitraum von einem Jahr gewählt. Diese Zykluslänge entspricht dem kleinsten Screeningintervall und ist aufgrund der langsamen Progression der intraepithelialen Neoplasien zum invasiven Karzinom als ausreichend klein anzusehen (Myers et al. 2000a; Mandelblatt J.S. et al. 2002).

Frauen konnten aufgrund eines invasiven Zervixkarzinoms basierend auf den krebsstadienspezifischen Überlebenskurven oder aus anderer Ursache in Abhängigkeit des Alters sterben, sie konnten in ihrem Gesundheitszustand verweilen oder zu einem anderen Gesundheitszustand progredieren oder regredieren. Die Entwicklung eines invasiven Zervixkarzinoms konnte nur durch eine Progressionsfolge über die zervikalen intraepithelialen Neoplasie Zustände CIN 1-3 erfolgen. Eine spontane Krebsentwicklung von "Gesund" nach "Zervixkarzinom FIGO I" (ohne vorherige Entwicklung von Dysplasien) innerhalb eines Jahres wurde im Modell nicht zugelassen und wird in der Fachliteratur auch als nicht wahrscheinlich angesehen. Das Modell lässt eine Entwicklung des Zervixkarzinoms von "Intraepitheliale Neoplasie Grad 3 (CIN 3)" direkt nach "Zervixkarzinom FIGO I" innerhalb eines Jahres zu. Die Karzinomentwicklung von einer "Intraepithelialen Neoplasie Grad 2 (CIN 2)" wurde indirekt über CIN 3 nach "Zervixkarzinom FIGO I" innerhalb von 2 Jahren oder mehr mit den entsprechenden Übergangswahrscheinlichkeiten im Modell berücksichtigt. Analog wurde die Karzinomentwicklung von einer "Intraepithelialen Neoplasie Grad 1 (1)" indirekt über CIN 2 und CIN 3 innerhalb von 3 Jahren oder mehr modelliert. Von den Zuständen CIN 1-3 war eine Regression zum Gesundheitszustand "Keine Läsion" möglich. Eine Regression vom Gesundheitszustand des invasiven Karzinoms wurde im Modell ausgeschlossen. Eine Frau konnte zu jeder Zeit mit Ausnahme der Stadien mit invasivem Karzinom eine ausschließlich benigne Hysterektomie erfahren und trat dann in den Zustand "Hysterektomie" ein, in dem kein spezifisches Screening für Zervixkarzinom mehr stattfand. Zervikale intraepitheliale Neoplasien (CIN 1-3) konnten durch ein auffälliges Screeningtestergebnis entdeckt und gemäß der deutschen klinischen Praxis kontrolliert oder behandelt werden. Behandelte Frauen gingen in den Gesundheitszustand "Keine Läsion" zurück. Unabhängig vom Screening konnte ein Zervixkarzinom aufgrund von klinischen Symptomen entdeckt werden. Wurde ein Zervixkarzinom entweder durch Screening oder aufgrund klinischer Symptomatik entdeckt, so wurde es behandelt. An Zervixkarzinom erkrankte und behandelte Frauen wurden nicht wieder in die Population für das primärpräventive Screening zurückgeführt. Es wurde angenommen, dass eine an Zervixkarzinom erkrankte und behandelte Frau nach dem 5. Jahr nach Erstdiagnose und des Therapiebeginns keine erhöhte Mortalität gegenüber Frauen ohne Zervixkarzinom mehr hatten und die Mortalität wieder der altersabhängigen Hintergrundmortalität gesunder Frauen entsprach.

Es wurde gemäß des deutschen Krebsfrüherkennungsprogramms ein Mindestalter für das Screening von 20 Jahren angesetzt. In der Basisfall-Analyse wurde ein einjähriges Screeningintervall gemäß des gesetzlich festgelegten Krebsfrüherkennungsprogramms in Deutschland festgelegt. In weiteren Szenarien wurden andere Mindestalter und die Screeningintervalle zwei, drei, und fünf Jahre analysiert.

Abbildung 11 veranschaulicht das Markov-Modell mit den möglichen Zuständen und Übergangsmöglichkeiten in Form eines vereinfachten Markov-Blasendiagramms (Bubble Diagram).

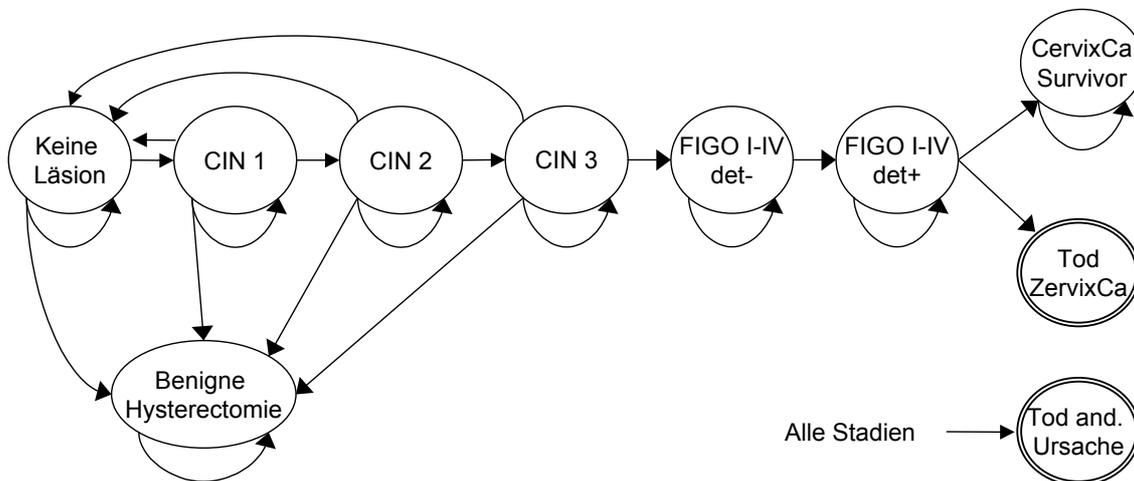


Abbildung 11: Markov-Modell in Form eines Bubble Diagram

CIN: Cervical Intraepithelial Neoplasia; det-: nicht detektiertes Zervixkarzinom; det+: detektiertes Zervixkarzinom; FIGO: Karzinomeinteilung nach Klassifikation der Fédération Internationale de Gynécologie et d'Obstétrique. CIN 3 repräsentiert CIN 3 plus Carcinoma in situ.

III C 4. 9 Parameter des Entscheidungsmodells

III C 4. 9. 1 Stadienspezifische Progression, Regression und Entdeckungsraten

Die jährlichen Übergangswahrscheinlichkeiten der zyto-histologischen Progressionen und Regressionen sowie die jährlichen Entdeckungswahrscheinlichkeiten aufgrund klinischer Symptome, die im Markov-Modell verwendet wurden, wurden in zwei Schritten festgelegt:

1. Das Modell wurde zunächst als Startkonfiguration mit epidemiologischen Parametern aus der internationalen Literatur besetzt.
2. Ausgehend von dieser Startkonfiguration wurden, wo dies aufgrund der vorliegenden Daten möglich war, die Übergangswahrscheinlichkeiten anhand von Be-

obachtungsdaten aus deutschen Krebsregistern an den deutschen Kontext angepasst.

Für den ersten Schritt wurden die Daten von möglichst qualitativ hochwertigen publizierten Studien herangezogen, die die derzeitige wissenschaftlichen Evidenz des Erkrankungsverlaufs widerspiegeln und dabei die deutsche zyto-histologische Klassifikation verwenden bzw. der Modellstruktur möglichst gut entsprechen. Der zweite Schritt stellt einen Datenanpassungsprozess dar und dient der Kalibrierung des Markov-Modelles. Dadurch wird sichergestellt, dass das Modell die deutschen Verhältnisse bezüglich der Epidemiologie und der klinischen Routinestandards repräsentiert. In dem nachfolgenden Unterkapitel wird dieses Vorgehen ausführlich beschrieben.

In Tabelle 49 sind die im entscheidungsanalytischen Modell verwendeten jährlichen Progressions- und Regressionswahrscheinlichkeiten aufgeführt und die Literaturstellen der internationalen Studien angegeben, auf denen diese Parameter beruhen.

Tabelle 49: Jährliche Übergangswahrscheinlichkeiten im Markov-Modell

Zustand vor Übergang	Zustand nach Übergang	Alter (Jahre)	Jährliche Wahrscheinlichkeit	Quelle
Keine Läsion	CIN 1	15-55+	0,0017-0,0521	(Kiviat 1996; Koutsky 1997; Ho et al. 1998; Hildesheim et al. 1994) in (Myers et al. 2000a; Mandelblatt J.S. et al. 2002)
CIN 1	CIN 2	15-34	0,0173	(Ho et al. 1998; Munoz et al. 1996; Oortmarssen & Habbema 1991; Syrjanen et al. 1992) in (Myers et al. 2000a)
		35+	0,0595	
CIN 2	CIN 3		0,0567	(Syrjanen et al. 1992; Kataja et al. 1992; Östör 1993)
CIN 1	Keine Läsion	15-34	0,1027	(Ho et al. 1998; Munoz et al. 1996; Oortmarssen & Habbema 1991; Syrjanen et al. 1992) in (Myers et al. 2000a)
		35+	0,0645	
CIN 2	Keine Läsion	15-34	0,1027	(Ho et al. 1998; Munoz et al. 1996; Oortmarssen & Habbema 1991; Syrjanen et al. 1992) in (Myers et al. 2000a)
		35+	0,0645	
CIN 3	Keine Läsion		0,0567	(Ho et al. 1998; Munoz et al. 1996; Oortmarssen & Habbema 1991; Syrjanen et al. 1992) in (Myers et al. 2000a)
CIN 3	FIGO I		0,0410	(Fahs et al. 1992; Eddy 1990; Myers et al. 2000a; Kim et al. 2002; Mandelblatt J.S. et al. 2002) *
FIGO I	FIGO II		0,2015	(Myers et al. 2000a)
FIGO II	FIGO III		0,2592	(Myers et al. 2000a)
FIGO III	FIGO IV		0,3624	(Myers et al. 2000a)
Nicht-Symptomatisch	Symptomatisch			
FIGO I	FIGO I		0,1098	(Myers et al. 2000a)*
FIGO II	FIGO II		0,2150	(Myers et al. 2000a)*
FIGO III	FIGO III		0,6120	(Myers et al. 2000a)*
FIGO IV	FIGO IV		0,9000	(Myers et al. 2000a)*

CIN: cervical intraepithelial Neoplasia; FIGO: Karzinomstadium nach Klassifikation der Fédération Internationale de Gynécologie et d'Obstétrique; Karzinoma in situ ist in CIN 3 enthalten. * Eigene Anpassung.

III C 4. 9. 2 Modellkalibrierung und Anpassung an den deutschen Kontext

Die folgenden Parameter wurden an deutsche Krebsregisterdaten für eine ungescreente Population aus den Jahren 1960 bis 1964, zur Verfügung gestellt durch das Gemeinsame Krebsregister Berlin, angepasst:

- die Progressionswahrscheinlichkeit von „Keine Läsion“ nach „CIN 1“
- die Progressionswahrscheinlichkeit von „CIN 3“ zum invasiven Karzinom „FIGO I“

- die Progressionswahrscheinlichkeiten innerhalb der FIGO-Stadien zum nächst schwereren Stadium
- die Entdeckungswahrscheinlichkeiten von invasiven Karzinomen „FIGO I bis IV“

Dabei wurden die für diesen Bereich üblichen statistischen Kalibrierungsverfahren verwendet, wie sie auch von anderen Autoren für die Kalibrierung publizierter internationaler Modelle für das Screening bei Zervixkarzinom eingesetzt wurden (Goldie S. et al. 2001; Myers et al. 2000a; Mandelblatt J.S. et al. 2002). Diese Verfahren machen sich folgende Eigenschaften zunutze:

1. Die Gesamtinzidenz der entdeckten Zervixkarzinomfälle in einer nicht gescreenten Population ist eine Funktion der Progressionswahrscheinlichkeit von CIN 3 zum invasiven Karzinom FIGO I und der symptomatischen Entdeckungswahrscheinlichkeit. Da die jährlichen Gesamtinzidenzen entdeckter Zervixkarzinomfälle aus Krebsregistern sowohl gesamt also auch altersspezifisch bekannt ist, können diese zum Anpassen der Progression von CIN 3 zu FIGO I und der symptomatischen Entdeckungswahrscheinlichkeit verwendet werden.
2. Die Verteilung der Zervixkarzinomstadien FIGO I bis IV in einer nicht gescreenten Population ist eine Funktion der Progressionswahrscheinlichkeit des Karzinoms und der stadienspezifischen symptomatischen Entdeckungswahrscheinlichkeit. Da die Verteilung der Zervixkarzinomstadien aus Registerdaten bekannt ist, können diese zum Anpassen der Progressionsraten und symptomatischen Entdeckungswahrscheinlichkeiten verwendet werden.

Die Beschreibung einer nicht gescreenten deutschen Population erfolgte anhand der epidemiologischen Krebsregisterdaten vor Einführung des gesetzlichen Krebsfrüherkennungsprogramms 1971. Es wurde auf die altersspezifischen und altersstandardisierten Zervixkarzinom-Inzidenzen insgesamt und stratifiziert nach FIGO-Stadium zurückgegriffen.

Das Krebsregister Saarland stellte den Autoren im Rahmen des HTA-Projektes epidemiologische Registerdaten ab 1968 zur Verfügung. Die Bereitstellung des Originaldatensatzes erfolgte durch Prof. Brenner (Ruprecht-Karls-Universität Heidelberg) (Brenner et al. 1990b). Vom Gemeinsamen Krebsregister (GKR) Berlin wurden entsprechende Daten von den neuen Bundesländern ab 1964 zur Verfügung gestellt. Die Einteilung in Stadien 1 bis 4 in den Gemeinsamen Krebsregistern Berlin erfolgte nicht basierend auf der TNM-Klassifikation, sondern nach klinischen Dokumentationen. Nach Angaben des Krebsregisters stellen diese jedoch eine gute Approximation

an die FIGO-Stadien dar (persönliche Kommunikation Dr. Stabenow, Gemeinsames Krebsregister (GKR), Berlin, 28.03.2002). Da vereinzelt ein Screening bereits in den Jahren vor der gesetzlichen Einführung des Krebsfrüherkennungsprogramms nicht auszuschließen ist, wurden für die Primäranalyse Daten der späten 60er Jahre der Gemeinsamen Krebsregister Berlin herangezogen. Die Stadienverteilung der Gemeinsamen Krebsregister Berlin in den späten 60er Jahren ist vergleichbar mit der Verteilung des Saarlands (Brenner et al. 1990b) in den späten 60er Jahren, so dass von einer robusten Datenlage ausgegangen werden kann.

Technisch wurden für den Anpassungsprozess der Modellparameter an die Beobachtungsdaten aus den Krebsregistern die Modellparameter als unabhängige Variablen (Progressions- und Entdeckungswahrscheinlichkeiten) eingesetzt und die Beobachtungsdaten der Krebsregister als abhängige Variablen (beobachtete Gesamtinzidenz der entdeckten Zervixkarzinomfälle und Verteilung der FIGO-Stadien). Die Modellparameter wurden im Iterationsverfahren solange variiert, bis simultan eine optimale Anpassung der vom Modell prädizierten Gesamt-Zervixkarzinominzidenz und FIGO-Verteilung an die aus den Krebsregistern ermittelte Gesamt-Zervixkarzinominzidenz und FIGO-Verteilung erreicht wurde.

Das Modell berücksichtigt eine altersabhängige Progression von dem Gesundheitszustand „Keine Läsion“ zu einer milden Dysplasie „CIN 1“. Unter Verwendung der Annahme, dass in den meisten Fällen eine HPV-Infektion eine notwendige Bedingung für diese Progression ist, konnten von Myers et al. 2000 publizierte Daten zur altersabhängigen HPV-Inzidenz in den Anpassungsprozess einbezogen werden (Myers et al. 2000a). Die Transformation der altersabhängigen HPV-Inzidenzen in die altersabhängigen Inzidenzen für CIN 1 erfolgte univariat mit einer relativen Progressionsrate. Ausgehend von den publizierten jährlichen HPV-Inzidenzraten als Starteinstellung wurden unter Beibehaltung der relativen Verhältnisse die Progressionsraten von „Keine Läsion“ zu „CIN 1“ so angepasst, dass die vom Modell prädizierte Zervixkarzinominzidenz im Bereich der beobachteten Werte für eine ungescreente Population in Deutschland lag.

Die Gesamtinzidenz des Zervixkarzinoms wurde außerdem zur Anpassung der Progressionswahrscheinlichkeit von CIN 3 zum invasiven Karzinom FIGO I verwendet. Die vom Modell für die Kohorte prädizierte kumulative Lebenszeitinzidenz des Zervixkarzinoms wurde durch die Restlebenserwartung dividiert, um die prädizierte jährliche Zervixkarzinominzidenz zu berechnen. Diese wurde an die aus den Registerdaten ermittelte, beobachtete altersstandardisierte Zervixkarzinominzidenz von 46,1 pro 100.000 pro Jahr durch Variation der Progressionswahrscheinlichkeit von CIN 3 nach FIGO I angepasst. Nachfolgend wurden die Progressionsraten der einzelnen FIGO-Zustände und die Wahrscheinlichkeiten für die zur Entdeckung des

Zervixkarzinoms führenden Symptome iterativ variiert, um die vom Modell prädizierte FIGO-Verteilung an die beobachteten Daten aus den Krebsregistern anzupassen. Die FIGO-Verteilung wurde berechnet aus den Anteilen der kumulativen FIGO-spezifischen Inzidenzen und der kumulativen Gesamtinzidenz. Tabelle 50 zeigt die Daten, die für den Anpassungsprozess von Progressionswahrscheinlichkeiten und Entdeckungswahrscheinlichkeiten zur Prädiktion einer ungescreenten deutschen Population verwendet wurden.

Tabelle 50: Verteilung der Zervixkarzinomstadien nach FIGO in Prozent. Quelle: Daten der Registerstelle des Gemeinsamen Krebsregisters (GKR) Berlin (Jahr 1964-66)

	Häufigkeit (%)
FIGO I	38.84
FIGO II	31.62
FIGO III	24.10
FIGO IV	5.44
Altersstandardisierte Inzidenz (pro 100.000)	
Zervixkarzinom	46,1

FIGO: Karzinomstadium nach Klassifikation der Fédération Internationale de Gynécologie et d'Obstétrique

Das Modell wurde einer internen und externen Validierung unterzogen. Die Ergebnisse der Validierung werden separat im Abschnitt III C 4. 10 beschrieben.

III C 4. 9. 3 Hysterektomieraten

Es lagen keine Daten zur altersspezifischen benignen Hysterektomierate in Deutschland in der Literatur vor. Aus diesem Grund wurden die Daten einer englischen Studie von Vasse^y et al. (Vessey & Villard-Mackintosh 1992) eingesetzt.

Das Modell prädizierte für Deutschland ein Lebenszeitrisko für eine Hysterektomie von 30,5%, welches gut mit den Angaben im Frauengesundheitsbericht von Bremen 2001 (Anonymous 2002) übereinstimmt (Einzelheiten siehe Kapitel III C 4. 10).

Tabelle 51: Altersspezifische jährliche benigne Hysterektomierate aufgrund anderer Ursachen als ein Zervixkarzinom. Quelle: (Vessey & Villard-Mackintosh 1992), UK (Daten von 1985-1989)

Alter	Jährliche Rate benigner Hysterektomie(%)
35-39	0,884
40-44	1,125
45-49	1,074
50 und älter	0,597

III C 4. 9. 4 Mortalität

Die FIGO-stadienspezifischen jährlichen Überlebenswahrscheinlichkeiten bei einem Zervixkarzinom im 1., 2., 3., 4., und 5. Jahr nach Diagnose/Therapiebeginn wurden aus den Überlebenskurven der im Tumorregister München erfassten Zervixkarzinompatientinnen aus der Region München im Zeitraum 1988-1997 berechnet (zur Verfügung gestellt durch das Tumorregister München, s. Tabelle 52). Die entsprechenden 5-Jahres-Überlebensraten für FIGO I bis IV waren 94,3%, 73,6%, 59,4% und 23,8%. Es wurde keine zervixkarzinombedingte Mortalität nach fünf Jahren nach Diagnose bzw. Therapiebeginn angenommen, da ein Versterben aufgrund von Zervixkrebs im Vergleich zu den ersten fünf Jahren relativ gering ist (Myers et al. 2000a). Dies deckte sich auch mit den Daten des Tumorregisters München.

Für die Modellierung der Hintergrundsterblichkeit wurden die altersspezifischen Sterberaten der deutschen weiblichen Allgemeinbevölkerung aus dem Statistischen Bundesamtes 1997-1999 herangezogen (Statistisches Bundesamt: Sterbetafel 1997-1999).

III C 4. 9. 5 Effektivitätsdaten

Es konnten weder zur Effektivität des Papanicolaou-Screeningtestverfahrens noch zu den neuen Technologien in der Literatur aktuelle deutsche Studien identifiziert werden. Daten zur Sensitivität und Spezifität des Papanicolaou-Testverfahrens und den neuen Technologien wurden deshalb aus internationalen Studien entnommen.

Tabelle 52: Jährliche Überlebenswahrscheinlichkeiten nach Zervixkarzinomstadien (FIGO) in Prozent. Quelle: Tumorregister München, Daten Region München von 1988-1997

Zustand	Jahr nach Diagnose	Jährliche Überlebenswahrscheinlichkeit (%)
FIGO I	1	98,6
	2	97,2
	3	97,9
	4	98,9
	5	97,6
FIGO II	1	92,8
	2	84,4
	3	91,4
	4	89,2
	5	95,9
FIGO III	1	84,4
	2	78,6
	3	82,5
	4	87,8
	5	91,3
FIGO IV	1	46,1
	2	62,7
	3	58,1
	4	85,7
	5	89,6

FIGO: Karzinomstadium nach Klassifikation der Fédération Internationale de Gynécologie et d'Obstétrique

Die Grundlage bildeten die im medizinischen und ökonomischen Teil dieses HTA-Reports eingeschlossenen internationalen HTA-Berichte (McCroy & Matchar 1999; AHTAC 1998; Broadstock 2000; Payne et al. 2000), die in diesen HTA-Berichten angegebenen Studien zu neuen Primärscreeningverfahren (ThinPrep, AutoCytePrep und AutoPap Primary Screen) sowie die jüngst veröffentlichten und im medizinischen Teil bewerteten Studien zu den Primärscreeningverfahren.

Für den Einsatz von Studienergebnissen als Parameter in einem entscheidungsanalytischen Screeningmodell müssen die Studien bzw. die berichteten Daten verschiedene Kriterien erfüllen, damit das Modell valide Ergebnisse für die Rahmenbedingungen der Evaluation generiert. Die folgenden Validitätskriterien, die bei der Auswahl der Daten für das Modell eine wichtige Rolle spielen, sind besonders hervorzuheben:

- Repräsentation der Zielpopulation der Fragestellung (beim Bevölkerungsscreening keine vorselektierte Risikogruppen)
- Geeignete Klassifikation der Krankheit bzw. der Krankheitsstadien
- Geeignete Klassifikation der Testergebnisse
- Valider Goldstandard in der diagnostischen Studie
- Einsatz des Goldstandards bei Testnegativen unabhängig von A-priori-Information (Vermeidung eines Verification Bias)

Im Folgenden soll detailliert auf diese Punkte eingegangen und die Selektionskriterien zu Studien mit Testgütedaten für die Verwendung im entscheidungsanalytischen Modell beschrieben werden. Die Beschreibung erfolgt getrennt für die Daten zum konventionellen Papanicolaou-Testverfahrens und den neuen, im Rahmen dieses HTAs bewerteten Testverfahrens.

Testcharakteristika des konventionellen Papanicolaou-Testverfahrens

In Deutschland wird die Münchner Zytologie-Klassifikation des Papanicolaou-Tests verwendet. Internationale Studien zur Sensitivität des Papanicolaou-Tests sind jedoch vorwiegend in Bethesda-Klassifikation (ASCUS, low-grade SIL, high-grade SIL) oder in der Histologie-Klassifikation nach Richart (CIN 1, CIN 2, CIN 3 und CIS) angegeben. Die Krankheitszustände (Markov-Zustände) des entscheidungsanalytischen Modells wurde aus diesem Grund basierend auf der CIN-Klassifikation aufgebaut, da diese leichter auf die Münchener Pap-Klassifikation abgebildet werden kann (CIN 1 und 2: Pap IIID; CIN 3 und CIS: Pap IV; Invasives Karzinom: Pap V). Die Bethesda-Klassifikation „ASCUS“ ist nicht übertragbar auf die Münchner Klassifikation, so dass Daten von Studien, die sich am zytologischen Schwellenwert „ASCUS“ orientierten, nicht berücksichtigt werden konnten.

Um bei der Selektion geeigneter Daten zur Sensitivität und Spezifität des konventionellen Papanicolaou-Testverfahrens für das entscheidungsanalytische Modell den obengenannten Validitätskriterien gerecht zu werden und die Zielpopulation dieses HTAs valide zu repräsentieren, wurde konkret auf die folgenden Punkte geachtet:

1. Es sollte sich bei den Studienpopulationen um Frauen mit durchschnittlichem Erkrankungsrisiko handeln und nicht um aufgrund von bereits stattgefundenen Tests, bestehender Symptomatik oder anderen Merkmalen vorselektierte Risikopopulationen.

2. Der Schwellenwert für ein testpositives Ergebnis des zytologischen Tests in der Studie sollte bei der Berechnung von Sensitivität und Spezifität unabhängig vom tatsächlichen Erkrankungsstadium auf geringgradige Läsionen (LSIL) festgelegt gewesen sein.
3. Nach Möglichkeit sollten statt einfacher Gesamt-Sensitivität und -Spezifität die Likelihoods für das Testergebnis "LSIL oder höher" stratifiziert nach dem jeweiligen tatsächlichen histologischen Gesundheitszustand (CIN 1, CIN 2, CIN 3/CIS, invasives Karzinom) berichtet worden sein.
4. Es sollte ein gebräuchlicher Goldstandard (Histologie bzw. unabhängige zweite zytologische Bewertung) bei positiven Befunden eingesetzt worden sein. Ferner sollte eine ausreichend große Fraktion von testnegativen Frauen mit dem Goldstandard untersucht worden sein, um eine valide Berechnung der Sensitivität unter allen testpositiven und testnegativen Kranken zu ermöglichen.
5. Die Auswahl der Patientinnen, bei denen der Goldstandard eingesetzt wurde, sollte zufällig und unabhängig von Risikofaktoren oder diagnostischen A-priori-Information erfolgt sein. Dies ist wichtig, um eine durch Verification Bias verzerrte Schätzung der tatsächlichen diagnostischen Gütemaße zu erhalten.

Die im medizinischen Berichtsteil dieses HTAs diskutierten Daten zur Sensitivität und Spezifität des konventionellen Papanicolaou-Verfahrens aus der Meta-Analyse von McCrory et al. (McCrory & Matchar 1999) konnten in dieser Form nicht im entscheidungsanalytischen Modell angewendet werden, da in dieser Studie (i) vorselektierte Studienpopulationen mit einem erhöhten Erkrankungsrisiko in die Meta-Analyse eingeschlossen wurden, (ii) teils der Schwellenwert „ASCUS“ verwendet wurde, und (iii) Studien mit Verification Bias enthalten waren. In einem weiteren Review derselben Forschergruppe (Nanda et al. 2000) wurden jedoch basierend auf dem Studienpool des systematischen Reviews von McCrory et al. (McCrory & Matchar 1999) Sensitivitäten und Spezifitäten disaggregiert für (i) Populationen mit geringem Erkrankungsrisiko, (ii) verschiedene Test-Schwellenwerte und (iii) um Verification Bias bereinigt dargestellt. Die Daten, welche die Validitätskriterien erfüllten wurden von den Autoren dieses HTAs in eigenen Berechnungen gepoolt und die gepoolten Parameter ins entscheidungsanalytische Modell übernommen (Tabelle 53).

Im englischen HTA von Payne et al. (Payne et al. 2000) wurden die nach dem tatsächlichen histologischen CIN-Zustand stratifizierten Likelihoods bezüglich der beiden Test-Schwellenwerten „milde Dysplasie“ und „moderate/schwere Dysplasie“

angegeben. Aus diesen Angaben wurde der jeweilige Anteil der Testergebnisse „milde Dysplasie“ und „moderate/schwere Dysplasie“ entnommen (Tabelle 53).

Tabelle 53: Sensitivität und Spezifität des Papanicolaou-Screeningtests (nach Nanda et al. 2000) und Anteile der milden und moderat-schwerwiegenden Testergebnisse (nach Payne et al. 2000) für die Stadien CIN 1, CIN 2, CIN 3 und invasives Karzinom.

Zustand	Spezifität in % (95%KI*)	Sensitivität in % (95%KI*)	Anteil Testergebnis Mild (%)	Anteil Testergebnis Moderat/Schwer (%)
Keine CIN	95,0 (94,5-96,4)		90,0	10,0
CIN 1		47,1 (44,8-49,4)	90,7	9,3
CIN 2		71,8 [†] (67,0-76,2)	59,5	40,5
CIN 3		71,8 [†] (67,0-76,2)	50,0	50,0
Invasives Karzinom		71,8 [†] (67,0-76,2)	60,0	40,0

95% KI: 95%-Konfidenzintervall

* Konfidenzintervalle von den Autoren berechnet

[†] Für CIN 2, CIN 3 und invasives Karzinom lagen die Daten zur Sensitivität nur zusammengefasst vor.

Testcharakteristika der neuen Primärscreeningtestverfahren

Die im Modell verwendeten Daten zur Sensitivität und Spezifität der in Deutschland auf dem Markt befindlichen neuen Primärscreeningverfahren (ThinPrep, AutoCyte-Prep und AutoPap Primary Screen) stammen aus den in den internationalen HTA-Berichten (McCrorry & Matchar 1999; AHTAC 1998; Broadstock 2000; Payne et al. 2000) eingeschlossenen Studien sowie aus den jüngst veröffentlichten und bereits im medizinischen Teil dieses HTA-Reports dargestellten und bewerteten Studien. Es wurden alle Studien, die einen Vergleich der neuen Verfahren mit dem konventionellen Papanicolaou-Test mit einem Testschwellenwert von "LSIL und höher" disaggregiert auf der Basis von absoluten Fallzahlen berichteten, eingeschlossen (Tabelle 54 bis Tabelle 57).

Für diese Daten wurde ein fallzahlgewichtetes Pooling durchgeführt und für den Vergleich des jeweiligen neuen Verfahrens versus konventionellen Pap-Test die folgenden Effektmaße für die Sensitivität berechnet:

- Absolute Differenzen der Sensitivität (Sens):
 $\Delta\text{Sens} = \text{Sens}(\text{Neues Verfahren}) - \text{Sens}(\text{Pap-Test})$
- Relative Falsch-Negativ-Rate (FNR):
 $\text{Relative FNR} = \text{FNR}(\text{Neues Verfahren}) / \text{FNR}(\text{Pap-Test})$
 $= (1 - \text{Sens}(\text{Neues Verfahren})) / (1 - \text{Sens}(\text{Pap-Test}))$
- Richtig-Positiv Odds Ratio (OR_{TP}) für eine Erhöhung der Richtig-Positiv-Odds (TP Odds) bzw. eine Erniedrigung der Falsch-Negativ-Odds (FN Odds):
 $\text{OR}_{\text{TP}} = \text{OR}_{\text{FN}} = \frac{[\text{Sens}(\text{Neues Verfahren}) / (1 - \text{Sens}(\text{Neues Verfahren}))]}{[(\text{Sens}(\text{Pap-Test}) / (1 - \text{Sens}(\text{Pap-Test})))]}$

Die rohen Daten und errechneten gepoolten Maße sind in Tabelle 54 bis Tabelle 57 dargestellt.

Im entscheidungsanalytischen Modell wurde schließlich die jeweilige Richtig-Positiv Odds Ratio (OR_{TP}) auf die Sensitivität des konventionellen Pap-Tests (Werte aus Nanda et al. 2000; siehe Tabelle 53) angewandt, um die Sensitivität der neuen Strategie zu erhalten:

$$\text{Sens}(\text{Neues Verfahren}) = \frac{[(\text{Sens}(\text{Pap-Test}) / (1 - \text{Sens}(\text{Pap-Test})) * \text{OR}_{\text{TP}}]}{\{1 + [(\text{Sens}(\text{Pap-Test}) / (1 - \text{Sens}(\text{Pap-Test})) * \text{OR}_{\text{TP}}]\}}$$

Dieser Ansatz minimiert potentielle Fehler aufgrund von Studienheterogenitäten.

Es lag nur eine Studie zur Effektivität der Dünnschichtpräparation mit anschließendem automatisierten Primärscreeningverfahren vor (Minge et al. 2000). In dieser wurde die Sensitivität nur für den Schwellenwert LSIL+/CIN1+ angegeben. Im entscheidungsanalytischen Modell entsprach die Sensitivität für den Schwellenwert LSIL+/CIN2+ dem Wert der Sensitivität einer Dünnschichtpräparation mit manueller Auswertung für den Schwellenwert LSIL+/CIN2+.

Von den oben genannten Studien berichten nur zwei (Bolick & Hellman 1998; Sheets et al. 1995) die Spezifität der Dünnschichtpräparation im Vergleich zum konventionellen Verfahren, jedoch nur für den Schwellenwert LSIL+/CIN 1+. Keine Studie berichtet die Spezifität für den Vergleich von automatisiertem Screening zum Konventionellem. Aus diesem Grund wurde wie auch in anderen internationalen Entscheidungsanalysen (McCrary & Matchar 1999; Hutchinson & Farber 2000; Payne et al. 2000) für die neuen Screeningverfahren dieselbe Spezifität wie für das konventionelle Papanicolaou-Verfahren eingesetzt.

Tabelle 54: Sensitivität der Dünnschichtpräparation. Schwellenwert LSIL+/CIN 1+.

Studie	Test verfahren	Konventioneller Pap-Test		Dünnschichtpräparation		Dünnschichtpräparation vs. Pap-Test		
		Testpositiv n	Goldstand- dard-positiv N	Testpositiv n	Goldstand- dard-positiv N	Δ Sens	Relative FNR	OR _{TP}
Schwellenwert LSIL+/CIN 1+								
(Sheets et al. 1995)	ThinPrep	107	159	117	159	0,063	0,808	1,354
(Corkill et al. 1997)	ThinPrep	29	84	60	84	0,369	0,436	4,741
(Sherman et al. 1998)	ThinPrep	374	549	443	549	0,126	0,606	1,956
(Bolick & Hellman 1998)	ThinPrep	57	67	40	42	0,102	0,319	3,509
(Hutchinson et al. 1999)	ThinPrep	210	323	257	323	0,146	0,584	2,095
(Weintraub & Morabia 2000)	ThinPrep	277	528	252	312	0,283	0,405	3,806
(Papillo et al. 1998)	ThinPrep	153	212	113	140	0,085	0,693	1,614
(Park et al. 2001)	ThinPrep	76	87	72	87	-0,046	1,364	0,695
(Bishop et al. 1998)	AutoCytePrep	73	93	83	93	0,108	0,500	2,274
(Data on file, CellPath, 1999)	AutoCytePrep	363	470	361	470	-0,004	1,019	0,976
(Vassilakos et al. 2000)	AutoCytePrep	124	140	690	758	0,025	0,785	1,309
(Minge et al. 2000)	AutoCytePrep	41	81	38	81	-0,037	1,075	0,862
Summe/gepoolter Wert (95% KI)		1884	2793	2526	3098	0,141	0,567	2,131 (1,89; 2,40)

Δ Sens Differenz Sensitivität, Relative FNR: relative Falsch-Negativ-Rate, OR_{TP}: Richtig-Positiv Odds Ratio, 95% KI: 95% Konfidenzintervall

Tabelle 55: Sensitivität der Dünnschichtpräparation. Schwellenwert LSIL+/CIN 2+.

Studie	Test verfahren	Konventioneller Pap-Test	Dünnschichtpräparation	Dünnschichtpräparation vs. Pap-Test	OR _{TP}			
		Test-positiv n	Goldstand-positiv N	Test-positiv n	Goldstand-positiv N	ΔSens	Relative FNR	OR _{TP}
<i>Schwellenwert LSIL+/CIN 2+</i>								
(Weintraub & Morabia 2000)	ThinPrep	175	339	159	189	0,325	0,328	4,967
(Hutchinson et al. 1999)	ThinPrep	103	137	73	79	0,172	0,306	4,016
(Inhorn 1998)	ThinPrep	44	47	45	47	0,021	0,667	1,534
(Park et al. 2001)	ThinPrep	65	72	63	72	-0,028	1,286	0,754
(Vassilakos et al. 2000)	AutoCytePrep	75	79	432	453	0,004	0,916	1,097
Summe/gepoolter Wert (95% KI)		387	595	340	387	0,228	0,347	3,888 (2,74; 5,51)

ΔSens Differenz Sensitivität, Relative FNR: relative Falsch-Negativ-Rate, OR_{TP}: Richtig-Positiv Odds Ratio, 95% KI: 95% Konfidenzintervall

Tabelle 56: Sensitivität der automatisierten Primärscreeningverfahren. Schwellenwert LSIL+/CIN 1+ und Schwellenwert LSIL+/CIN 2+.

Studie	Test verfahren	Konventionelles Screening		Automatisiertes Screening		Automatisiertes vs. Konventionelles Screening		OR _{TP}	
		Test-positiv n	Goldstandard-positiv N	Test-positiv n	Goldstandard-positiv N	ΔSens	Relative FNR		
<i>Schwellenwert LSIL+/CIN 1+</i>									
(Wilbur et al. 1998)	AutoPap Primärscreen	298	348	321	348	0,066	0,540	1,995	
Summe/gepoolter Wert (95% KI)		298	348	321	348	0,066	0,540	1,995 (1,22; 3,27)	
<i>Schwellenwert LSIL+/CIN 2+</i>									
(Wilbur et al. 1999)	AutoPap Primärscreen	65	70	68	70	0,043	0,400	2,615	
Summe/gepoolter Wert (95% KI)		65	70	68	70	0,043	0,400	2,615 (0,49; 13,96)	

ΔSens Differenz Sensitivität, Relative FNR: relative Falsch-Negativ-Rate, OR_{TP}: Richtig-Positiv Odds Ratio, 95% KI: 95% Konfidenzintervall

Tabelle 57: Sensitivität der Dünnschichtpräparation mit anschließendem automatisierten Primärscreeningverfahren. Schwellenwert LSIL+/CIN 1+.

Studie	Test verfahren	Konventioneller Pap-Test		Neue Screeningverfahren		Neues vs. Konventionelles Screening	OR _{TP}
		Test-positiv n	Goldstand- dard-positiv N	Test-positiv n	Goldstand- dard-positiv N		
Schwellenwert LSIL+/CIN 1+							
(Minge et al. 2000)	AutoCytePrep + AutoCyteScreen	41	81	59	81	0,222	0,550
Summe/gepoolter Wert (95% KI)		41	81	59	81	0,222	0,550
							2,616 (1,36; 5,04)

ΔSens Differenz Sensitivität, Relative FNR: relative Falsch-Negativ-Rate, OR_{TP}: Richtig-Positiv Odds Ratio, 95% KI: 95% Konfidenzintervall

III C 4. 9. 6 Kosten

Die Erhebung von ökonomischen Primärdaten wie Mengen- und Preisgerüste und ihre Zusammenführung zu Kostendaten für das Zervixkarzinom-Screening und die Zervixkarzinomtherapie in Deutschland wurde von den ökonomischen Experten der HTA-Arbeitsgruppe (Prof. Dr. Wasem, Dr. Aidelsburger) in Abstimmung mit den weiteren Arbeitsgruppenmitgliedern durchgeführt und bereits im Abschnitt II C 4.7 „Erhebung ökonomischer Primärdaten für Deutschland“ ausführlich beschrieben.

Als Kostenarten wurden ausschließlich direkte medizinische Kosten wie Screeningtests, confirmatorische diagnostische Tests, ambulante und stationäre Versorgung, Arzneimitteltherapie, operative Therapie und palliative Maßnahmen berücksichtigt. Quellen für die Mengengerüste waren vorwiegend die Empfehlungen zur Diagnostik, Therapie und Nachsorge des Tumorzentrums München (Anonymous 1999b), die durch Expertenschätzungen ergänzt wurden. Wo erforderlich, wurde auf administrative Preise zurückgegriffen. Stationäre Kosten wurden über die Verweildauern kombiniert mit Pflegesätzen und Sonderentgelten unter zusätzlicher Berücksichtigung von PKV-spezifischen Kosten erfasst und bewertet. Leistungen der ambulanten Versorgung wurden basierend auf der Gebührenordnung für Ärzte (GOÄ) für Privatversicherte und dem Einheitlichen Bewertungsmaßstab (EBM) für gesetzliche Versicherte bewertet. Das Basisjahr der Erhebung war das Jahr 2000. Alle Kosten wurden mit einer jährlichen Rate von 3% in der Entscheidungsanalyse diskontiert.

Die Kosten des konventionellen Screeningverfahren wurde basierend auf den Abrechnungsziffern GOÄ Nr. 27 und EBM Nr. 157 für die gynäkologische Untersuchung und Abstrichentnahme gemäß der gesetzlichen Früherkennungsuntersuchung sowie den Ziffern GOÄ Nr. 4851 und EBM Nr. 4951 für die Färbung und zytologische Bewertung des Abstrichs berechnet.

In der Berechnung der Kosten der neuen Screeningverfahren wurden die Kosten für die gynäkologische Untersuchung einschließlich Abstrichentnahme sowie die Kosten für die Färbung und zytologische Bewertung übernommen und aus der Mikrokostenberechnung die zusätzlichen Kosten für die Durchführung der neuen Verfahren, bestehend aus Materialkosten sowie Kapital-, und Instandhaltungskosten unter Berücksichtigung von Abschreibungen und einer jährlichen Verzinsung von 3%, hinzu addiert. Bei diesem Ansatz werden Unterschiede in den Personalkosten, die zum Beispiel durch eine schnellere Auswertung des Präparates durch das neue Verfahren im Vergleich zum konventionellen Verfahren entstehen können, nicht

berücksichtigt. Es wurde ein gewichteter Mittelwert aus den zusätzlichen Kosten der beiden Dünnschichtpräparationsverfahren ThinPrep und AutoCytePrep berechnet und für die Dünnschichtpräparationsverfahren eingesetzt.

Tabelle 58 gibt eine Übersicht über die im Modell verwendeten Kosten. Es werden undiskontierte Kosten pro Leistung bzw. für Leistungen über den Zeitraum von einem Jahr angegeben.

Tabelle 58: Undiskontierte Kosten für Screening, Diagnostik, Behandlung und Palliativmaßnahmen des Zervixkarzinoms

Leistung	Kosten (€)
Konventioneller Papanicolaou-Test und manuelle Auswertung	17,28
Dünnschichtpräparation und manuelle Auswertung	29,03
Konventioneller Papanicolaou-Test und automatisierte Auswertung	18,41
Dünnschichtpräparation und automatisierte Auswertung	30,16
Beratungsgespräch und gynäkologische Untersuchung	8,07
Kolposkopie mit Knipsbiopsie	11,48
Konisation mit Diagnostik und histologischer Auswertung	604,79
Diagnostik bei Symptomentwicklung	394,28
Behandlung bei Karzinom FIGO IA	2000,11
Behandlung bei Karzinom FIGO IB / IIAB	6381,58
Behandlung bei Karzinom FIGO III	9050,56
Behandlung bei Karzinom FIGO IV	7883,14
Behandlungskosten bei Endstadium	8798,82
Jährliche Arzneimittelkosten bei FIGO IB / IIAB	215,68
Jährliche Arzneimittelkosten bei FIGO III	748,28
Jährliche Arzneimittelkosten bei FIGO IV	749,22
Jährliche Nachbehandlungskosten bis zu 2 Jahre nach Therapie	2007,39
Jährliche Nachbehandlungskosten 3 bis 5 Jahre nach Therapie	502,81
Jährliche Nachbehandlungskosten ab 6 Jahre nach Therapie	48,70

III C 4. 10 Validierung des Modells

Das Modell und seine Parameter wurden in zwei Schritten validiert:

1. Interne Validierung durch einen Vergleich der Modellprädiktionen mit den epidemiologischen Daten aus dem Gesamtkrebsregister Berlin.
2. Externe Validierung anhand (i) veröffentlichter deutscher Daten aus der Literatur und (ii) der epidemiologischen Daten aus dem Krebsregister Saarland.

Die Ergebnisse dieser beiden Schritte für die verschiedenen Zielgrößen werden im Folgenden wiedergegeben.

III C 4. 10. 1 Zervixkarzinominzidenz

Nach den Daten des Gemeinsamen Krebsregister Berlin liegt für eine ungescreente Population der beobachtete Altersgipfel für die Zervixkarzinominzidenz in der Altersgruppe 55-59 Jahre mit einer Peak-Zervixkarzinominzidenz von 112/100.000 Frauen. Das entscheidungsanalytische Modell prädizierte bei einem Altersgipfel von 51 Jahren eine Peak-Zervixkarzinominzidenz von 84/100.000 Frauen pro Jahr. Gustafsson et al. (Gustafsson et al. 1997) berichteten für ungescreente Populationen in verschiedenen Ländern altersspezifische Zervixkarzinom-Inzidenzen, die ähnliche Muster ergaben. In Deutschland lag in den Jahren 1964-1968 der Altersgipfel bei 47,5 Jahren mit einer jährlichen Peak-Zervixkarzinominzidenz von 106,6/100.000 Frauen in Hamburg und 112,3/100.000 Frauen in den neuen Bundesländern. Aus den Daten des Krebsregisters Saarland ist für die Jahre 1968-1970 ein Altersgipfel der Zervixkarzinominzidenz von 53 Jahren zu entnehmen. Damit liegen die vom Modell prädizierten Daten sowohl für die jährliche Gesamtinzidenz als auch für die Altersgipfel mit den zugehörigen altersspezifischen Inzidenzen im Bereich der tatsächlich in Deutschland beobachteten Daten.

III C 4. 10. 2 Verteilung der FIGO-Stadien

Die Gesamt-FIGO-Verteilung des Modelles konnte sehr präzise an die in dem Gemeinsamen Krebsregister Berlin beobachteten Daten angepasst werden. Das Modell prädizierte, dass 38,8% der entdeckten Zervixkarzinomfälle im Stadium FIGO I

entdeckt werden, 31,6% im Stadium FIGO II, 24,0% im Stadium FIGO III und 5,5% im Stadium FIGO IV. Dies entsprach der beobachteten Verteilung des Gemeinsamen Krebsregister Berlin. Die externen, nicht im Anpassungsprozess verwendeten Daten des Krebsregisters Saarland ergeben für die Jahre 1968-1970 eine Verteilung von ca. 38% für FIGO I, 27% für FIGO II und 35% für FIGO III+IV. Damit ergibt sich für die externe Validierung eine Abweichung der absoluten Prozentwerte zwischen 0,8-5,5%.

III C 4. 10. 3 Prävalenzen der CIN-Stadien

Das Modell prädizierte eine mittlere Gesamtprävalenz CIN 1-3 und CIS von 7,2%. Der Altersgipfel der Dysplasien lag bei 27 Jahren. Schneider et al. (Schneider A. et al. 2000) untersuchten eine deutsche Population von 967 Frauen in Thüringen, die weder eine Zervixdysplasie noch einen abnormalen zytologischen Testbefund innerhalb eines Jahres vor Studienbeginn zeigten. Es zeigte sich eine CIN 2/CIN 3-Prävalenz von 6,8% bei Frauen jünger als 35 Jahre und 1,5% in älteren Frauen mit einem Altersgipfel in der Gruppe der 20-35 Jährigen, welches eine rohe, nicht altersstandardisierte Gesamtprävalenz für CIN 2/CIN 3 von 3,9% ergab. Das entscheidungsanalytische Modell prädizierte eine Gesamtprävalenz für CIN 2/CIN 3 von 2,6%, welche aufgrund der Kohortensimulation für die gesamte deutsche weibliche Kohorte altersstandardisiert ist.

III C 4. 10. 4 Lebenszeitrisiko Zervixkarzinom

Das Lebenszeitrisiko einer Frau, an einem invasiven Zervixkarzinom zu erkranken, wurde vom Modell auf 3,0% prädiziert. Dies liegt innerhalb der Spannweite anderer internationaler Publikationen von 2,5-3,3% (Kim et al. 2002; McCrory & Matchar 1999; Myers et al. 2000a; Brown & Garber 1999). Daten des Krebsregisters Saarland für den Zeitraum 1968-1970 ergaben ein Risiko von 3,3%, bis zum 85. Jahr an Zervixkarzinom zu erkranken (Brenner et al. 1990b).

III C 4. 10. 5 Lebenszeitrisiko für benigne Hysterektomie

Das vom Modell prädizierte Lebenszeitrisiko einer Frau für eine benigne Hysterektomie aus anderer Ursache als Zervixkrebs lag bei 30,5%. Im Frauengesundheitsbe-

richt Bremen 2001 (Anonymous 2002) wird angegeben, dass etwa jede dritte Frau im Laufe ihres Lebens von einer Hysterektomie betroffen ist und dass nur etwa 10% der Hysterektomien aufgrund eines Zervixkarzinoms durchgeführt wurden. Hieraus ergibt sich ein Lebenszeitrisiko einer Frau, eine benigne Hysterektomie aus anderer Ursache als Zervixkrebs zu erfahren, von etwa 30%, was gut mit der Modellprädiktion übereinstimmt. Tabelle 59 gibt eine Übersicht der Modellprädiktionen im Vergleich zu den beobachteten epidemiologischen Daten aus Deutschland. Ferner wurden zu Vergleichszwecken die Ergebnisse der Modellprädiktionen aus dem entscheidungsanalytischen Modell von Myers et al. (Myers et al. 2000a) gegenübergestellt. Die Gruppe Myers et al. ist Urheber des internationalen entscheidungsanalytischen Modells der im Rahmen des HTA-Berichtes von McCrory et al. 1999 durchgeführten Kosten-Effektivitätsstudie. Die medizinischen und gesundheitsökonomischen Studiergebnisse wurden sowohl im medizinischen Teil als auch im ökonomischen Teil dieses HTA-Berichts ausführlich diskutiert.

Tabelle 59: Vergleich der Modellprädiktion mit Beobachtungsdaten für eine ungescreente deutsche Population (interne und externe Modellvalidierung)

	Prädiktion	Interne Validierung	Externe Validierung	Vergleich internat. Literatur
	Deutsches Modell	Beobachtete Daten ¹	Beobachtete Daten ²⁻⁴	Myers et al. 2000
Altersgipfel CIN-Stadium (Jahre)				
CIN 1	25			28
CIN 2	38			42 [†]
CIN 3	48			
Altersgipfel (Jahre) Zervixkarzinom	51	55-59	47 ³ , 53 ²	48
Peak-Inzidenz Zervixkarzinom (pro 100.000)	84	112	107-112 ³	81
Gesamtinzidenz Zervixkarzinom (pro 100.000)	45,9	46,1		
Verteilung der FIGO-Stadien (%)				
FIGO I	38,8	38,8	38 ²	46,4
FIGO II	31,6	31,6	27 ²	27,0
FIGO III	24,0	24,1	35 ^{2*}	18,1
FIGO IV	5,5	5,4		8,5
Lebenszeitrisiko Zervixkarzinom (%)	3,0		3,3 ² (bis 85 J.)	2,5

* FIGO III+IV gemeinsam

† CIN 2/3 gemeinsam

Quelle deutscher Beobachtungsdaten: ¹Gemeinsames Krebsregister Berlin (Daten von 1964-1968), ²Krebsregister Saarland (Daten von 1968-1970) / Publikation Brenner et al. 1990, ³Gustafsson et al. 1997.

III C 4. 11 Modellannahmen

Dem entscheidungsanalytischen Modell wurden verschiedene Annahmen zugrunde gelegt, die im Folgenden aufgeführt werden.

Klassifikation

- Angaben in der Bethesda Klassifikation wurden wie folgt zugeordnet: Low-Grade Squamous Intraepithelial Lesion (LG-SIL oder LSIL) in CIN 1, High-Grade Squamous Intraepithelial Lesion (HG-SIL oder HSIL) in CIN 2, CIN 3 und Karzinom in situ.
- Die Klassifikationsgruppe Atypical Squamous Cells of Undetermined Significance (ASCUS) ist nicht in die Münchner Nomenklatur zu übertragen und ist ein zytologischer und kein pathologischer Befund. ASCUS-Befunde wurden als negative Testbefunde gewertet. Alle im Modell verwendeten Werte für Sensitivität und Spezifität repräsentieren den zytologischen Grenzwert LSIL und höher.

Progression/Regression

- Das Startalter der Markov-Kohorte wurde auf 15 Jahre festgelegt. Es wurde angenommen, dass bis zu einem Alter von 15 Jahren keine intraepithelialen Läsionen und keine invasiven Zervixkarzinome auftreten.
- Der Zustand CIN 1 repräsentiert eine Kombination der natürlichen Erkrankungszustände „HPV-Infektion“ und „CIN 1“. Die Wahl dieser Modellstruktur basierte (1) auf den in der Literatur berichteten neuesten wissenschaftlichen Erkenntnissen, dass eine HPV-Infektion wahrscheinlich in den meisten Fällen eine Voraussetzung für die Entwicklung einer intraepithelialen Läsion darstellt bzw. die Entwicklung einer intraepithelialen Läsion ohne eine HPV-Infektion selten vorkommt und (2) darauf dass sowohl die Progression von „HPV-Infektion“ zu „CIN 1“ als auch die Regression zu „Gesund“ in einem sehr kurzen Zeitraum und sehr häufig stattfindet (Schiffman et al. 1993; Kiviat 1996; Koutsky 1997), so dass nur wenige valide Daten zu diesen Progressions- und Regressionsraten existieren.
- Eine Zervixkarzinomentwicklung kann nur ausgehend von CIN 3 im Anschluss einer Progressionsabfolge von CIN 1 über CIN 2 nach CIN 3 erfolgen. Eine schnelle Progression zum invasiven Karzinom ausgehend von CIN 1 oder CIN 2 innerhalb von weniger als 2-3 Jahren wurde nicht berücksichtigt.

- Eine Regression der CIN-Stadien wurde jeweils nur zum Gesundheitszustand "Keine Läsion" zugelassen. Eine schrittweise Regression von CIN 3 nach CIN 2 und von CIN 2 nach CIN 1 wurde nicht berücksichtigt, da anzunehmen ist, dass die Regression sehr schnell verläuft und eine schrittweise Regressionsfolge bei einer Zyklusdauer von 1 Jahr in dem Parameter für die direkte Regression zum Zustand "Keine Läsion" enthalten ist.

Behandlung

- Entdeckte Läsionen werden entsprechend den deutschen Empfehlungen behandelt. Milde Läsionen (CIN 1) werden bei Persistenz über einen Zeitraum von mehr als 24 Monaten, moderate Läsionen (CIN 2) nach 12 Monaten oder bei Progression nach CIN 3 behandelt. Schwere Läsionen (CIN 3) werden sofort behandelt.
- Behandelte Läsionen, bei denen kein invasives Zervixkarzinom vorliegt, gelten als geheilt. Im Modell erfolgt ein Übergang der entsprechenden Patientinnen in den Gesundheitszustand "Keine Läsion". Damit besitzen diese bereits behandelten Frauen dasselbe CIN- und Zervixkarzinomrisiko wie Frauen, die bislang nicht behandelt wurden.
- Patientinnen mit Zervixkarzinom werden gemäß der deutschen Empfehlung entsprechend ihrem FIGO-Stadium behandelt.

Entdeckung des Zervixkarzinoms

- An Zervixkrebs erkrankte Frauen, die nicht am Screening teilnehmen bzw. bei denen ein falsch-negativer Testbefund bestand, können klinische Symptome entwickeln und hierdurch als Krebsfall entdeckt werden.
- Unabhängig von dem zytologischen Testbefund wurde angenommen, dass ein Zervixkarzinom bei einer Frau, die am Screening teilnimmt, allein durch die gynäkologische Untersuchung entdeckt werden kann. Diese Entdeckungsrate wurde basierend auf Expertenschätzungen für die Stadien FIGO I, FIGO II, FIGO III und FIGO IV auf 70%, 75%, 80% und 85% festgelegt.

Mortalität

- An Zervixkarzinom erkrankte und behandelte Patientinnen haben fünf Jahre nach der Diagnose bzw. ihrem Therapiebeginn gegenüber gesunden Frauen keine erhöhte Mortalität mehr, sondern es gilt die allgemeine altersspezifische Hintergrundmortalität der deutschen weiblichen Allgemeinbevölkerung.

Effektivität der Testverfahren

- Aufgrund der mangelhaften Datenlage zur Spezifität der Testverfahren wurden für die verschiedenen Testverfahren keine Unterschiede in der Spezifität berücksichtigt.
- Es wurde eine 100%ige Sensitivität für die Kombination von Kolposkopie mit Biopsie bei einer histologischen Abklärung einer intraepithelialen Läsion angenommen.

III C 4. 12 Szenarioanalysen, Subgruppenanalysen und Sensitivitätsanalysen

Neben der Basisfall-Analyse wurden folgende weitere Analysearten durchgeführt:

- Szenarioanalysen: Verwendung anderer "Screening Policies", wie beispielsweise andere Screeningintervalle oder Altersgrenzen für das Screening. Diese Analysen dienen zur Erweiterung der Erkenntnisse auf gesundheitspolitische Varianten des aktuell in Deutschland durchgeführten Screeningprogrammes.
- Subgruppenanalysen: Analyse von Subpopulationen wie bestimmter Altersgruppen. Diese Analysen dienen zur Identifikation von Subgruppen, für die differentielle Effektivität, Kosten oder Kosten-Effektivität bestehen. Subgruppenanalysen können wertvolle Hinweise für die Erstellung von Screening-Leitlinien geben.
- Sensitivitätsanalysen: Variation von einzelnen oder mehreren Modellparametern wie z.B. Testkosten oder Testsensitivität. Sensitivitätsanalysen dienen der Überprüfung der Robustheit der Ergebnisse der Basisfall-Analyse bzw. der Abschätzung der Unsicherheit der Ergebnisse. Ferner dienen sie der Identifikation unsicherer Parameter, für die weiterer Forschungsbedarf besteht.

III C 4. 12. 1 Szenarioanalysen

Screeningintervall

Es wurden neben der Basisfall-Analyse mit einem Screeningintervall von einem Jahr weitere Szenarien mit unterschiedlichen Screeningintervallen von 2, 3 und 5 Jahren analysiert. Zur besseren Übersicht und Interpretierbarkeit wurden die Ergebnisse der Analysen für die verschiedenen Screeningintervalle zusammen mit den Ergebnissen der Basisfall-Analyse (jährliches Screeningintervall) dargestellt.

Screening-Altersgrenze

Die untere Altersgrenze zum Einschluss in das Screeningprogramm wurde zwischen 15 und 35 Jahren variiert.

III C 4. 12. 2 Subgruppenanalysen

Das Anfangsalter der Kohorte wurde zwischen 15 und 75 Jahre variiert, um den Einfluss des Einsatzes der neuen Screeningverfahren in älteren Altersgruppen zu untersuchen.

III C 4. 12. 3 Sensitivitätsanalysen

Modellparameter

Ein-Weg-Sensitivitätsanalysen wurden unter Berücksichtigung der in der Literatur publizierten Spannweiten bzw. 95%-Konfidenzintervalle für alle Modellparameter durchgeführt. Einzelkosten wurden halbiert und verdoppelt. Die Diskontrate wurde zwischen 0% (keine Diskontierung) und 10% variiert. Die Werte für die Sensitivität der Screening-Testverfahren wurden basierend auf den zugehörigen 95%-Konfidenzintervallen der gepoolten Schätzer variiert. Die Spezifität der neuen Screeningverfahren wurde zwischen 80% und 100% variiert. Die Spezifität des konventionellen Screeningverfahren wurde basierend auf dem zugehörigen 95%-Konfidenzintervall des gepoolten Schätzers variiert. In einer Zwei-Weg-

Sensitivitätsanalyse wurden simultan die Kosten des Screeningtests und die Sensitivität der neuen Screening-Verfahren variiert.

Screening-Teilnahmeverhalten

Eine Modellierung des Einflusses von Screening-Teilnahmeverhalten auf den Effekt des Screeningprogrammes war nicht Gegenstand dieses HTA-Berichtes. Deshalb beziehen sich die Ergebnisse der Basisfall-Analyse durchgehend auf Frauen, die regelmäßig am Screeningprogramm teilnehmen, was numerisch eine Teilnahmerate von 100% am jährlichen Screening impliziert. Da jedoch dem Screening-Teilnahmeverhalten im Rahmen der Evaluation der Effektivität von Screeningprogrammen eine besondere Bedeutung zukommen kann und es zu den Aufgaben dieses HTAs gehört, weiteren Forschungsbedarf zu identifizieren, wurde im Sinne einer Pilotevaluation eine einfache Analyse der Teilnahmerate mit in die Evaluation aufgenommen. Dabei wurde die Teilnahmerate am Screening zwischen 0% (keine Teilnahme der Frauen am Screening) bis 100% (vollständige Teilnahme aller Frauen am Screening) variiert. Es wurde ausschließlich eine zufällige und von anderen Faktoren unabhängige Teilnahmerate pro Screeningzyklus verwendet, d.h. für jedes Jahr in dem ein Screening vorgesehen war, wurde erneut und unabhängig von einer früheren Screeningteilnahme mittels einer zufälligen Teilnahmerate die Teilnahme oder Nichtteilnahme einer Frau am Screening zum aktuellen Zeitpunkt simuliert. Für eine umfangreiche und realistischere Modellierung des Teilnahmeverhaltens wären weitere Daten bezüglich des Screeningverhaltens der deutschen Zielpopulation erforderlich, die nicht in ausreichendem Maße vorliegen. Aus den genannten Gründen kommt der Analyse des Teilnahmeverhaltens nur qualitative Bedeutung zu, die Aufschluss darüber geben soll, ob dem Teilnahmeverhalten beim Screening für Zervixkarzinom in Deutschland eine relevante Rolle zukommt oder nicht.

III C 4. 13 Software

Entscheidungsanalytische Berechnungen wurden mit DATA Pro for Health Care (TreeAge Software Inc., Williamstown, MA) durchgeführt. Alle weiteren statistischen Analysen wurden mit SAS 8.1 (SAS Institute Inc., Cary, NC) und Excel (Microsoft) durchgeführt. Für epidemiologische Berechnungen zu aggregierten Maßzahlen wurde zusätzlich die Software EpiCalc 2000 Version 1.02 (Gilman & Myatt 1998) eingesetzt.

III C 5 Ergebnisse

Im Folgenden werden die Ergebnisse der Entscheidungsanalyse dargestellt. Im Abschnitt III C 5. 1 werden zunächst die Ergebnisse zu den klinischen Endpunkten und der Lebenserwartung beschrieben, die sich für die verschiedenen untersuchten Screeningtechnologien im Vergleich ergeben. In Abschnitt III C 5. 2 werden die Langzeitkosten und die Kosten-Effektivität der einzelnen Screeningverfahren berichtet. Die Abschnitte III C 5. 4 und III C 5. 5 stellen schließlich die Ergebnisse der Subgruppen- und Sensitivitätsanalysen dar.

III C 5. 1 Effektivität

Die Basisfall-Analyse geht von einer Kohorte von Frauen im Alter von 15 Jahren aus, die ab dem Alter von 20 Jahren jährlich auf Zervixkarzinom gescreent werden, regelmäßig an diesem Screening teilnehmen (Teilnahmerate 100%) und über einen lebenslangen Zeithorizont betrachtet werden.

Die Basisfall-Analyse ergab für die Strategie "Kein Screening" eine Lebenserwartung von 65,80 Jahren, Die verschiedenen Screeningstrategien erzielten jeweils eine Lebenserwartung von rund 66,06 Jahre. Die Lebenserwartung einer gesunden Frau im Alter von 15 Jahren liegt nach den Sterbetafeln des Statistischen Bundesamtes bei 66,06 Jahren (Statistisches Bundesamt: Sterbetafel 1997-1999). Vergleicht man die verschiedenen untersuchten Screeningverfahren untereinander bezüglich der Lebenserwartung, liegen die Unterschiede über alle Screeningintervalle hinweg (1-5 Jahre) im Bereich von Bruchteilen von Tagen bis Wochen. Aus diesem Grunde werden im Folgenden diese Ergebnisse zusätzlich als gewonnene Lebenstage angegeben.

Konventionelles Screening mit der Papanicolaou-Methode und manueller Auswertung erhöht die Lebenserwartung gegenüber "Kein Screening" um 94,17 Tage. Die Papanicolaou-Methode mit anschließender automatisierter Auswertung erhöht die Lebenserwartung gegenüber "Kein Screening" um 94,62 Tage, die Dünnschichtpräparation mit manueller Auswertung um 94,67 Tage und die Kombination von Dünnschichtpräparation und automatisierter Auswertung um 94,67 Tage (Tabelle 60).

Die Modellrechnung ergab, dass in einer Kohorte von 100.000 Frauen ohne Screening 3032 Frauen in ihrer Lebenszeit an Zervixkarzinom erkranken und diagnostiziert werden. Mit konventionellem Screening sind dies nur noch 38 von 100.000 Frauen. Mit den neuen Screeningverfahren sinkt diese Zahl auf 6/100.000 Frauen bei einer automatisierten Auswertung des Papanicolaou-Test, 4/100.000 Frauen bei einer Dünnschichtpräparation plus manueller Auswertung und 3/100.000 Frauen bei einer Kombination von Dünnschichtpräparation und automatisierter Auswertung. (Tabelle 61). Mit anderen Worten verhindert das konventionelle Screening 2.994 Krebsfälle und die neuen Screeningverfahren zwischen 3026 und 3029 Krebsfälle pro 100.000 Frauen. Ein Umstieg von dem konventionellen auf eine der neuen Strategien würde zwischen 32 und 35 weitere Krebsfälle pro 100.000 Frauen verhindern.

Gegenüber der Strategie kein Screening verhindert konventionelles Screening 1000 zervixkarzinombedingte Todesfälle pro 100.000 Frauen. Mit den neuen Screeningverfahren würden den Modellrechnungen zufolge zusätzlich zwischen 3,3 und 3,6 Todesfälle pro 100.000 Frauen verhindert werden (Tabelle 62).

In Tabelle 60 ist die inkrementelle Lebenserwartung für die untersuchten Strategien bei verschiedenen Screeningintervallen dargestellt. Diese Ergebnisse zeigen, dass mit zunehmender Länge des Screeningintervalls die inkrementelle Lebenserwartung (gewonnene Lebensstage) aller Screeningstrategien im Vergleich zu keinem Screening abnimmt und damit die Effektivität des Screening bezüglich der Lebenserwartung mit der Häufigkeit zunimmt. Dieser Effekt bleibt auch bis zu einem jährlichen Screeningintervall erhalten. So lässt sich bereits mit dem konventionellen Verfahren die durchschnittliche Lebenserwartung um 94 zusätzliche Tage erhöhen, während beim 5-Jahresintervall nur 79 Tage gewonnen werden. Im Gegensatz dazu nimmt die inkrementelle Effektivität bezüglich der Lebenserwartung der neuen Screeningverfahren im Vergleich zum konventionellen Screening mit zunehmender Screeninghäufigkeit ab und ist unter den untersuchten Szenarien am geringsten bei einem Screeningintervall von einem Jahr. Während sich mit den neuen Screeningverfahren gegenüber dem konventionellen Verfahren bei einem 5-Jahresintervall noch 8-10 zusätzliche Tage Lebenserwartung gewinnen lassen, liegt dieser Wert für das 1-Jahresintervall unter einem Tag.

Dasselbe screeningintervallbedingte Muster ist bei der Betrachtung der Zervixkarzinom-Morbidität (bzw. verhinderte Zervixkarzinomfälle) und zervixkarzinom-spezifischen Mortalität (bzw. verhinderte zervixkarzinombedingte Todesfälle) zu beobachten (Tabelle 61 bzw. Tabelle 62). Während die Zahl der berechneten Krebsfälle beim 5-Jahresintervall je nach Screeningverfahren zwischen 252 und 699 pro 100.000 Frauen liegt, kann diese Zahl durch ein einjähriges Screeningintervall je nach Verfahren auf 3-38 Krebsfälle pro 100.000 Frauen gesenkt werden. Während

allerdings ausgehend vom jährlichen konventionellen Screening durch einen Wechsel zu einem jährlichen Screening mit einem der neuen Verfahren pro 100.000 Frauen maximal 35 Krebsfälle zusätzlich vermieden werden können, können beim 5-Jahresintervall bis zu 447 zusätzliche Krebsfälle vermieden werden. Analog würden bei einer Substitution des konventionellen durch eines der neuen Screeningverfahren beim 1-Jahresintervall pro 100.000 Frauen ca. 3 bis 4 krebsbedingte Todesfälle zusätzlich vermieden, während dies beim 5-Jahresintervall bereits zwischen 73 und 88 zusätzlich vermiedene Todesfälle wären. Tabelle 61 und Tabelle 62 verdeutlichen auch in absoluten Zahlen die Abnahme der Krebsfälle und Krebstodesfälle mit abnehmendem Screeningintervall.

Tabelle 60: Inkrementelle undiskontierte Lebenserwartung für die untersuchten Strategien bei verschiedenen Screeningintervallen.

Strategie	Screening-intervall (Jahre)	Inkrementelle Lebenserwartung vs. Kein Screening (Lebenstage)	Inkrementelle Lebenserwartung vs. Konventionelles Verfahren (Lebenstage)
Kein Screening	1	-	-
Pap + ManScreen	1	94,17	-
Pap + AutoScreen	1	94,62	0,46
DS + ManScreen	1	94,67	0,50
DS + AutoScreen	1	94,67	0,50
Kein Screening	2	-	-
Pap + ManScreen	2	91,68	-
Pap + AutoScreen	2	93,85	2,17
DS + ManScreen	2	94,14	2,46
DS + AutoScreen	2	94,18	2,50
Kein Screening	3	-	-
Pap + ManScreen	3	88,99	-
Pap + AutoScreen	3	92,75	3,77
DS + ManScreen	3	93,34	4,35
DS + AutoScreen	3	93,40	4,42
Kein Screening	5	-	-
Pap + ManScreen	5	79,48	-
Pap + AutoScreen	5	87,81	8,33
DS + ManScreen	5	89,44	9,96
DS + AutoScreen	5	89,60	10,11

Pap: Papanicolaou-Test, ManualScreen: manuelle Auswertung, AutoScreen: automatisierte Auswertung, DS: Dünnschichtpräparation

Tabelle 61: Absolute und inkrementelle verhinderte Krebsfälle pro 100.000 Frauen für die untersuchten Strategien bei verschiedenen Screeningintervallen.

Strategie	Screeningintervall (Jahre)	Absolute Krebsfälle pro 100.000	Verhinderte Krebsfälle pro 100.000 vs. Kein Screening	Verhinderte Krebsfälle pro 100.000 vs. Konventionelles Verfahren
Kein Screening	1	3032	-	-
Pap + ManScreen	1	38	2994	-
Pap + AutoScreen	1	6	3026	32
DS + ManScreen	1	4	3028	34
DS + AutoScreen	1	3	3029	35
Kein Screening	2	3032	-	-
Pap + ManScreen	2	180	2851	-
Pap + AutoScreen	2	53	2979	127
DS + ManScreen	2	36	2996	144
DS + AutoScreen	2	33	2998	147
Kein Screening	3	3032	-	-
Pap + ManScreen	3	310	2722	-
Pap + AutoScreen	3	112	2920	198
DS + ManScreen	3	81	2951	229
DS + AutoScreen	3	76	2955	234
Kein Screening	5	3032	-	-
Pap + ManScreen	5	699	2333	-
Pap + AutoScreen	5	333	2699	366
DS + ManScreen	5	261	2771	438
DS + AutoScreen	5	252	2780	447

Pap: Papanicolaou-Test, ManualScreen: manuelle Auswertung, AutoScreen: automatisierte Auswertung, DS: Dünnschichtpräparation

Tabelle 62: Absolute und inkrementelle verhinderte Krebstodesfälle pro 100.000 Frauen für die untersuchten Strategien bei verschiedenen Screeningintervallen.

	Screeningintervall (Jahre)	Absolute Krebs-Todesfälle pro 100.000	Verhinderte Krebs-Todesfälle pro 100.000 vs. Kein Screening	Verhinderte Krebs-Todesfälle pro 100.000 vs. Konventionelles Verfahren
Kein Screening	1	1004,1	-	-
Pap + ManScreen	1	3,9	1000,2	-
Pap + AutoScreen	1	0,6	1003,5	3,2
DS + ManScreen	1	0,4	1003,7	3,5
DS + AutoScreen	1	0,3	1003,8	3,5
Kein Screening	2	1004,1	-	-
Pap + ManScreen	2	22,7	981,4	-
Pap + AutoScreen	2	6,1	998,0	16,6
DS + ManScreen	2	3,9	1000,2	18,7
DS + AutoScreen	2	3,7	1000,4	18,9
Kein Screening	3	1004,1	-	-
Pap + ManScreen	3	44,7	959,4	-
Pap + AutoScreen	3	14,5	989,6	30,3
DS + ManScreen	3	10,0	994,1	34,5
DS + AutoScreen	3	9,5	994,6	35,0
Kein Screening	5	1004,1	-	-
Pap + ManScreen	5	127,9	876,2	-
Pap + AutoScreen	5	54,4	949,7	73,5
DS + ManScreen	5	40,8	963,3	86,5
DS + AutoScreen	5	39,4	964,7	87,8

Pap: Papanicolaou-Test, ManualScreen: manuelle Auswertung, AutoScreen: automatisierte Auswertung, DS: Dünnschichtpräparation

III C 5. 2 Kosten und Kosten-Effektivität

Tabelle 63 stellt die absoluten (undiskontierten) und diskontierten Kosten und Effekte sowie das inkrementelle diskontierte Kosten-Effektivitäts-Verhältnis für die untersuchten Strategien in der Basisfall-Analyse dar, die von einer Kohorte von Frauen im Alter von 15 Jahren, die ab einem Alter von 20 Jahren jährlich auf Zervixkarzinom gescreent werden, ausgeht.

Ohne Screening liegt der aus dem Modell berechnete Erwartungswert der undiskontierten absoluten Lebenszeitkosten bei 630 € pro Frau. Bei jährlichem Screening mit konventioneller Methode sind dies 1.220 €. Diese Lebenszeitkosten erhöhen sich bei Einsatz der neuen Screeningtechnologien im jährlichen Setting und liegen für die automatisierte Auswertung bei 1.350 €, für die Dünnschichtpräparation plus manuelle Auswertung bei 1.970 € und für die Dünnschichtpräparation plus automatisierte Auswertung bei 2.070 €. Die Lebenszeitkosten pro Frau erhöhen sich damit bei einem Screening mit den neuen Verfahren im Vergleich zum konventionellen Verfahren je nach Verfahren zwischen 130 und 840 € bzw. zwischen 11% und 69%. Nach Diskontierung ergeben sich bei der Strategie kein Screening Lebenszeitkosten von 190 €. Bei konventionellem Screening sind dies 550 €, bei einem automatisiertem Auswertungsverfahren 630 €, bei Dünnschichtpräparation mit manueller Auswertung 910 € und bei einer automatisierten Auswertung der Dünnschichtpräparation 950 €.

Im Vergleich zu keinem Screening hat ein jährliches Screening mit der konventionellen Papanicolaou-Methode ein diskontiertes inkrementelles Kosten-Effektivitäts-Verhältnis (IKEV) von 6.600 € pro gewonnenem Lebensjahr (€/LJ). Im Vergleich zu diesem konventionellen Screening beträgt für das jährliche Screening das diskontierte IKEV für die computergestützte Auswertung des Papanicolaou-Tests 220.000 €/LJ, für die Dünnschichtpräparation plus manuelle Auswertung 953.000 €/LJ und eine Kombination von Dünnschichtpräparation und computergestützter 1.083.000 €/LJ.

Basierend auf einem Schwellenwert der gesellschaftlichen Zahlungsbereitschaft von etwa 50.000 € pro Lebensjahr ist ein jährliches Screening mit der konventionellen Papanicolaou-Methode im Vergleich zu keinem Screening als kosteneffektiv anzusehen. Eine Einführung einer der neuen Screeningverfahren oder eine Kombination von Dünnschichtpräparation mit computergestützter Auswertung ist bei einem jährlichen Screening als nicht kosteneffektiv zu bewerten.

Tabelle 63: Basisfall-Analyse: Absolute Kosten und Effektivität sowie inkrementelle diskontierte Kosten-Effektivitäts-Verhältnisse für die untersuchten Strategien bei einem jährlichen Screeningintervall (gerundete Werte).

	Kein Screening	Pap + ManScreen	Pap + AutoScreen	DS + ManScreen	DS + AutoScreen
		vs. Kein Screening	vs. konventionelles Verfahren (Pap + ManScreen)		
Kosten (€)*	633	1.221	1.351	1.973	2.065
Lebenserwartung (Jahre)*	65,80	66,06	66,06	66,06	66,06
Gewonnene Lebenstage vs. kein Screening*	-	94,17	94,62	94,67	94,67
Diskontierte Kosten (€) [†]	186	553	628	906	954
Diskontierte Lebenserwartung (Jahre) [†]	28,57	28,63	28,63	28,63	28,63
Diskontierte gewonnene Lebenstage vs. kein Screening [†]	-	20,30	20,42	20,43	20,43
Diskontiertes IKEV (€/gewonnenem Lebensjahr) [†]	-	6.603 ¹	219.647 ²	952.568 ²	1.082.838 ²

Pap: Papanicolaou-Test, ManualScreen: manuelle Auswertung, AutoScreen: automatisierte Auswertung, DS: Dünnschichtpräparation

* Undiskontierte Werte

[†] 3% jährliche Diskontrate. Inkrementelle Kosten-Effektivitäts-Verhältnisse (IKEV) konventionelles Screening mit Papanicolaou-Test und manueller Auswertung versus keinem Screening¹, alle anderen Strategien jeweils zum konventionellem Screening².

In Abbildung 12 sind die undiskontierten absoluten Lebenszeitkosten eines Screenings mit den verschiedenen Technologien als Balkendiagramm aufgetragen. Diesen ist die im Vergleich zu keinem Screening gewonnene Lebenserwartung (in Lebenstagen) bei Einsatz der verschiedenen Screeningtechnologien als Kurvenzug gegenübergestellt.

Tabelle 64 zeigt die absoluten undiskontierten Kosten und Effekte sowie das diskontierte inkrementelle Kosten-Effektivitäts-Verhältnis für die untersuchten Strategien bei verschiedenen Screeningintervallen. Die Analyseergebnisse zeigen, dass mit zunehmendem Screeningintervall absolute undiskontierte Lebenszeitkosten und undiskontierte Lebenserwartung sowohl mit konventionellem Verfahren als auch mit den neuen Screeningtechnologien reduziert werden. So beträgt das gerundete inkrementelle diskontierte Kosten-Effektivitäts-Verhältnis von konventionellem Screening im Vergleich zu keinem Screening bei einem 1-Jahresintervall 6.600 €/LJ und reduziert

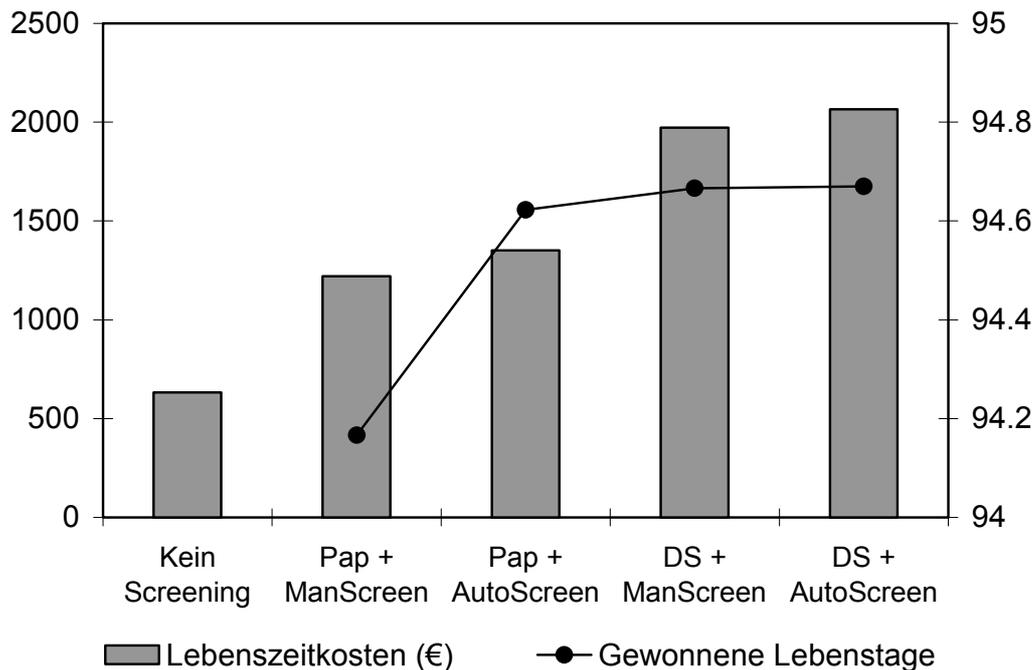


Abbildung 12: Absolute Lebenszeitkosten der untersuchten Screeningverfahren (in €) und inkrementelle gewonnene Lebensstage versus keinem Screening.

sich bei einem 2-Jahres-Screeningintervall auf 2.300 €/LJ, bei einem 3-Jahres-Screeningintervall auf 1.400 €/LJ und bei einem 5-Jahres-Screeningintervall auf 140 €/LJ. Analog verbessert sich die Kosten-Effektivität der neuen Screeningverfahren mit zunehmendem Screeningintervall.

Im Vergleich zum konventionellen Screeningverfahren reduziert sich das inkrementelle diskontierte Kosten-Effektivitäts-Verhältnis einer Kombination der neuen Screeningverfahren bei einem 2-Jahres-Screeningintervall von 1.083.000 €/LJ (bei 1-Jahresintervall) auf 119.000 €/LJ, bei einem 3-Jahres-Screeningintervall auf 53.000 €/LJ und bei einem 5-Jahres-Screeningintervall auf 13.000 €/LJ. Die inkrementellen diskontierten Kosten-Effektivitäts-Verhältnisse einer Einführung der computergestützten Auswertung oder der Dünnschichtpräparation liegen jeweils zwischen den Werten der konventionellen Papanicolaou-Methode und der Kombination der neuen Technologien.

Betrachtet man das Screening bei Screeningintervallen von 2 oder 3 Jahren, so ergeben sich unter Zugrundelegung einer gesellschaftlichen Zahlungsbereitschaft in der Größenordnung von 50.000 €/LJ kosteneffektive Bedingungen für die computergestützte Auswertung des Papanicolaou-Tests (für Screeningintervall ≥ 2 Jahre), für

die Dünnschichtpräparation plus manuelle Auswertung (für Screeningintervall ≥ 3 Jahre) und für die Kombination von Dünnschichtpräparation und computergestützter Auswertung (für Screeningintervall ≥ 3 Jahre). Es sei hier angemerkt, dass es sich bei diesen Ergebnissen durchweg um inkrementelle Vergleiche der neuen Screeningverfahren mit dem konventionellen Papanicolaou-Verfahren handelt und sich alle inkrementellen Kosten-Effektivitäts-Verhältnisse auf den Vergleich mit dem konventionellen Papanicolaou-Verfahren beziehen. Für robuste inkrementelle Vergleiche zwischen den drei evaluierten neuen Screeningtechnologien ist zum einen die Datelage unzureichend und zum anderen waren Vergleiche zwischen den einzelnen Strategien nicht Inhalt dieses HTAs. Es sei jedoch angemerkt, dass sich nach dem Prinzip des abnehmenden Grenznutzens häufig die Kosten-Effektivität einer noch effektiveren und noch teureren Technologie verschlechtert, wenn diese Anstelle mit dem Standardverfahren mit einer Technologie verglichen wird, die ihrerseits effektiver, teurer, aber kosteneffektiv (d.h. akzeptables IKEV) im Vergleich zum Standardverfahren ist.

Abbildung 13 zeigt die Effizienz-Grenzlinien (*cost-effectiveness frontiers*) für das konventionelle und die drei neuen Screeningtechnologien jeweils für die Screeningintervalle von 1, 2, 3 und 5 Jahren. Auf der Hochachse sind die diskontierten Lebenszeitkosten für die Screeningstrategien abgetragen und auf der Rechtsachse die dadurch erzielte medizinische Effektivität in diskontierten gewonnenen Lebenstagen.

Die diskontierten inkrementellen Kosten-Effektivitäts-Verhältnisse zwischen zwei Strategien werden in der Grafik getrennt für jedes Screeningintervall-Szenario durch die Steigung der Verbindungslinien zwischen den Symbolen für diese Strategien veranschaulicht. So repräsentiert in Abbildung 13 beispielsweise die schwarze Verbindungslinie zwischen dem Quadrat-Symbol und dem Dreiecks-Symbol die IKEV für den Vergleich computergestützte Auswertung des Papanicolaou-Tests (Pap+AutoScreen) versus konventionelles Screening (Pap+ManScreen) im Setting des jährlichen Screenings, d.h. 220.000 €/LJ.

Tabelle 64: Szenarien für verschiedene Screeningintervalle: Absolute undiskontierte Kosten und Effektivität sowie diskontierte inkrementelle Kosten-Effektivitäts-Verhältnisse für die untersuchten Strategien bei verschiedenen Screeningintervallen.

Strategie	Screening-intervall (Jahre)	Undiskontierte Kosten (€)	Undiskontierte Effektivität vs. Kein Screening (gewonnene Lebensstage)	Diskontiertes inkrementelles KE-Verhältnis (€/Lebensjahr)
Kein Screening	1	633	-	-
Pap + ManScreen	1	1.221	94,17	6.603 ¹
Pap + AutoScreen	1	1.351	94,62	219.647 ²
DS + ManScreen	1	1.973	94,67	952.568 ²
DS + AutoScreen	1	2.065	94,67	1.082.838 ²
Kein Screening	2	633	-	-
Pap + ManScreen	2	686	91,68	2.269 ¹
Pap + AutoScreen	2	749	93,85	26.438 ²
DS + ManScreen	2	1.064	94,14	103.434 ²
DS + AutoScreen	2	1.118	94,18	118.514 ²
Kein Screening	3	568	-	-
Pap + ManScreen	3	605	88,99	1.346 ¹
Pap + AutoScreen	3	633	92,75	11.455 ²
DS + ManScreen	3	843	93,34	46.279 ²
DS + AutoScreen	3	886	93,40	53.327 ²
Kein Screening	5	428	-	-
Pap + ManScreen	5	443	79,48	135 ¹
Pap + AutoScreen	5	555	87,81	1.501 ²
DS + ManScreen	5	581	89,44	10.810 ²
DS + AutoScreen	5	633	89,60	12.895 ²

Pap: Papanicolaou-Test, ManualScreen: manuelle Auswertung, AutoScreen: automatisierte Auswertung, DS: Dünnschichtpräparation

*3% jährliche Diskontrate. Inkrementelle Kosten-Effektivitäts-Verhältnisse konventionelles Screening mit Papanicolaou-Test und manueller Auswertung versus keinem Screening¹, alle anderen Strategien jeweils zum konventionellem Screening².

In Abbildung 13 deutlich sichtbar ist die Reduktion von absoluten Kosten und medizinischer Effektivität mit zunehmendem Screeningintervall (Vergleich der 4 verschiedenfarbigen Linien), was sich dadurch ausdrückt, dass die Kurve für das 5-Jahresintervall (rote Kurve) am weitesten unten und links im Schaubild liegt. Ferner lässt sich aus diesem Schaubild ablesen, dass die IKEV mit zunehmender Effektivität der Screeningverfahren zunehmen, dass also weitere Gewinne an Lebenserwartung immer teurer "erkauf" werden müssen. Die Zunahme der Steigungen von links nach rechts veranschaulicht das Prinzip des abnehmenden Grenznutzens, d.h. die Tatsa-

che, dass für jede weitere Einheit an Input von Ressourcen (Testkosten und medizinische Versorgung) der dafür erzielbare Output (Gewinn an Lebenserwartung) abnimmt.

Abbildung 13 veranschaulicht ferner das bereits oben dargestellte Ergebnis, dass die IKEV für die neuen versus dem konventionellen Screeningverfahren mit zunehmender Länge des Screeningintervalls immer günstiger werden und für das Szenario des jährlichen Screenings sehr steil (nicht kosteneffektiv) ist. Wie oben beschrieben (siehe auch Tabelle 64), wäre beispielsweise das Screening mit einer computergestützten Auswertung des Papanicolaou-Tests erst bei einem 2-Jahresintervall kosteneffektiv (IKEV ca. 26.000 €/LJ), wenn man diese Strategie mit dem konventionellen Verfahren vergleicht.

Die besondere Stärke des Schaubildes mit der Effizienzlinie (Abbildung 13) liegt in der Möglichkeit, simultan Vergleiche zu verschiedenen Screeningtechnologien und verschiedenen Screeningintervallen anzustellen. Dies erlaubt die vergleichende Bewertung aller 16 Screeningstrategien, die sich aus der Kombination der 4 untersuchten Screeningtechnologien und der 4 untersuchten Screeningintervalle ergeben. Screeningstrategien, die im Schaubild nicht auf der Effizienz-Grenzlinie, d.h. der konvexen Hüllkurve zu liegen kommen, sondern innerhalb dieser liegen, werden durch andere Screeningverfahren oder deren Kombinationen dominiert. So werden beispielsweise die beiden Screeningstrategien mit Dünnschichtverfahren (DS+ManScreen und DS+AutoScreen) im 3-Jahresintervall (Abbildung 13, Kreis und Raute auf blauer Linie) durch die Strategie mit computergestützter Auswertung des Papanicolaou-Tests im 2-Jahresintervall (Abbildung 13, Dreieck auf blauer Linie) dominiert, weil letztere bei geringeren Gesamtkosten eine höhere Screening-Effektivität besitzt.

So lässt sich als wichtige Erkenntnis aus diesem Schaubild ableiten, dass für jede der drei neuen Technologien ein Screening mit 2-Jahresintervall fast so effektiv ist wie konventionelles jährliches Screening, aber zum Teil deutlich weniger Kosten verursacht.

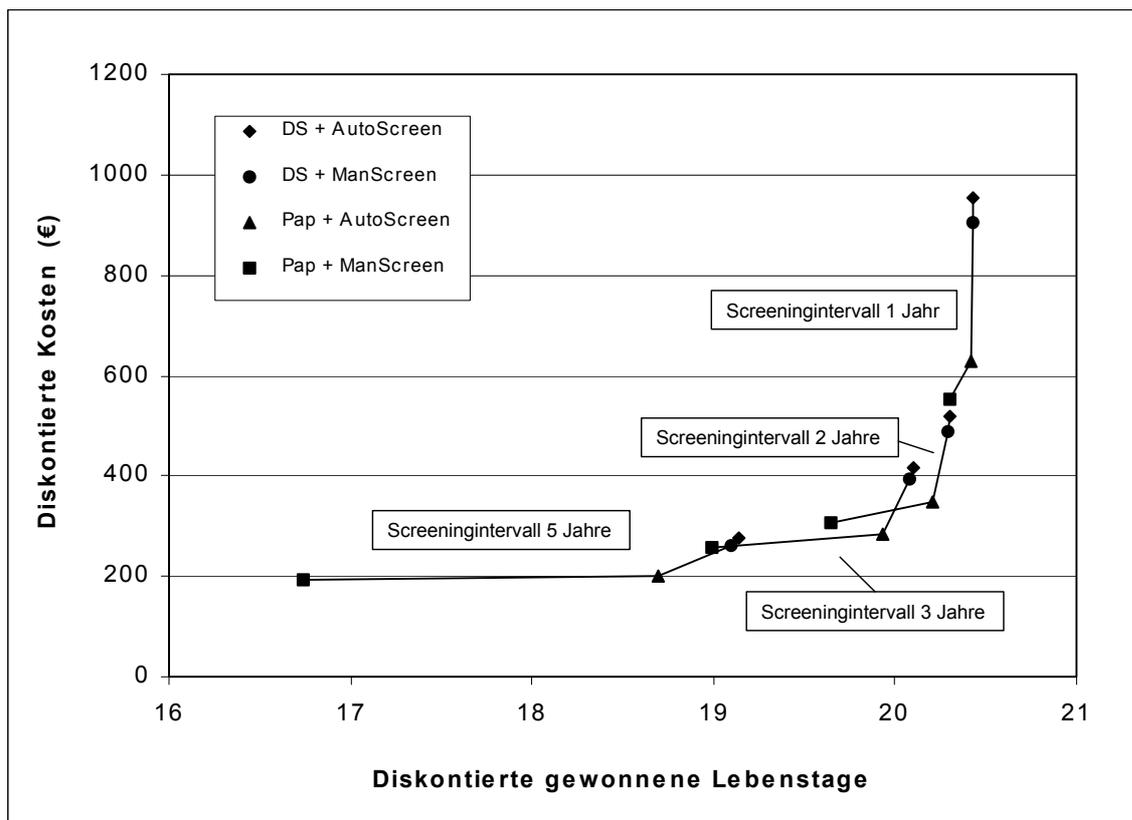


Abbildung 13: Szenario: Kosten-Effektivitäts-Frontier für die untersuchten Strategien bei verschiedenen Screeningintervallen (Diskontierte Kosten und Effekte)

Pap: Papanicolaou-Test, ManualScreen: manuelle Auswertung, AutoScreen: automatisierte Auswertung, DS: Dünnschichtpräparation

III C 5. 3 Szenarioanalysen

Die Szenarioanalysen zu den verschiedenen Screeningintervallen wurden zur besseren Übersicht bereits in den vorangehenden Abschnitten III C 5. 1 und III C 5. 2 zusammen mit den Ergebnissen der Basisfall-Analyse dargestellt.

Es folgt eine Beschreibung der Ergebnisse der Szenarioanalyse zum Alter als Einschlusskriterium in das Zervixkarzinom-Screeningprogramm. Die untere Altersgrenze zum Einschluss in das Screeningprogramm wurde zwischen 15 und 65 Jahren variiert, der Zeithorizont der Analyse (ab 15 Jahre bis Tod) wurde beibehalten. In Abbildung 14 werden die inkrementellen Kosten-Effektivitäts-Verhältnisse der verschiedenen Strategien bei Variation der unteren Screening-Altersgrenze dargestellt. Mit höherer unterer Screening-Altersgrenze werden die IKEV zunächst günstiger, steigen jedoch bei einem Screeningbeginn erst in höherem Alter wieder an. Das

konventionelle Screening ist für die gesamte Spannweite der unteren Altersgrenze kosteneffektiv, wenn man sich auf die Betrachtung eines einjährigen Screeningintervalles bei regelmäßiger Teilnahme (Teilnahmerate 100%) beschränkt. Bei Screeningbeginn im Alter von ca. ≥ 35 Jahren liegt die IKEV des Papanicolaou-Verfahrens mit automatisierter Auswertung unterhalb von 50.000 €/LJ und kann als kosteneffektiv bewertet werden. Es sei darauf hingewiesen, dass dieses Szenario voraussetzt, dass vor Erreichen der unteren Screening-Altersgrenze kein Screening stattfand. Die Ergebnisse haben also keine Gültigkeit für Frauen, die beispielsweise zwar über 35 Jahre sind, aber bereits an einem Screeningprogramm teilgenommen haben. Für eine solche Fragestellung sei auf die Ergebnisse der Subgruppenanalyse für Alter im Abschnitt III C 5. 4 verwiesen.

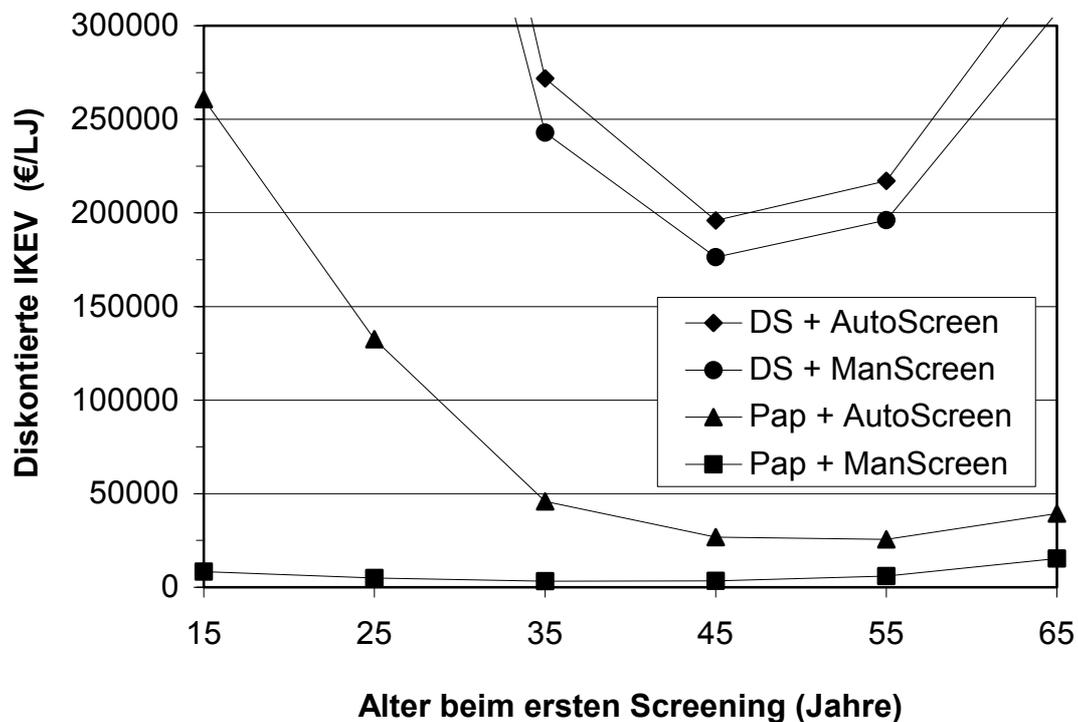


Abbildung 14: Einweg-Sensitivitätsanalyse: Inkrementelle Kosten-Effektivitäts-Verhältnisse (in €/Lebensjahr) der untersuchten Strategien bei unterschiedlicher unterer Screening-Altersgrenze.

Pap: Papanicolaou-Test, ManualScreen: manuelle Auswertung, AutoScreen: automatisierte Auswertung, DS: Dünnschichtpräparation

III C 5. 4 Subgruppenanalyse

In der Basisfall-Analyse wurde von einer Kohorte von Frauen im Alter von 15 Jahren ausgegangen. Würde ein neues Screeningprogramm eingeführt werden, so wären davon auch Frauen betroffen, die bei der erstmaligen Durchführung des neuen Screeningprogrammes bereits älter wären. Aus diesem Grunde wurden getrennte Analysen für verschiedene Altersgruppen durchgeführt.

Die Variation des Alters führte zu keinen vom Basisfall abweichenden Entscheidungen. Für alle drei untersuchten neuen Screeningtechnologien stiegen die inkrementellen Kosten-Effektivitäts-Verhältnisse mit zunehmendem Alter an und lagen damit bei allen drei neuen Screeningtechnologien über den Werten, die sich für diese Screeningstrategien in der Basis-Analyse für 15-jährige Frauen ergaben (siehe Tabelle 64). Dies bedeutet, dass für ein jährliches Screening die als nicht kosteneffektiv bewerteten neuen Screeningstrategien auch für ältere Frauen keine kosteneffektive Option darstellen. Dieser Sachverhalt galt für alle drei evaluierten neuen Screeningtechnologien

III C 5. 5 Sensitivitätsanalysen

III C 5. 5. 1 Modellparameter

Für die im Methodenteil genannten Parameter wurden Ein-Weg- bzw. Zwei-Weg-Sensitivitätsanalysen ausgehend vom Basisfall (Zeithorizont: 15 Jahre bis Tod, jährliches Screening ab dem 21. Lebensjahr) durchgeführt.

Die qualitativen Aussagen der Modellrechnungen blieben im Vergleich zur Basisfall-Analyse über die Spannweite der Parametervariation stabil. Variationen von Diskontierung, Kosten der Behandlungen, zusätzliche Kosten der neuen Screeningverfahren, als auch Progressions- und Regressionsraten der CIN-Zustände in Ein-Weg-Sensitivitätsanalysen hatten keinen Effekt auf den Entscheidungsausgang. Die Variation der Werte für die Sensitivität als auch für die Spezifität sowohl des konventionellen Verfahrens als auch der neuen Screeningverfahren innerhalb ihrer Konfidenzintervalle hatte ebenfalls keine Auswirkung auf die Entscheidung. Eine Ausnahme bildete die als zufällige Wahrscheinlichkeit modellierte Teilnehmerate am Screening, welche einen deutlichen Einfluss auf die Modellergebnisse aufwies und im Folgenden beschrieben wird.

III C 5. 5. 2 Screening-Teilnahmeverhalten

Die Basisfall-Analyse wurde durchgeführt für Frauen, die regelmäßig am Screeningprogramm teilnehmen, was implizit einer Teilnahmerate von 100% am jährlichen Screening entspricht. Eine detaillierte Evaluation des Einflusses des Screening-Teilnahmeverhaltens war nicht Ziel dieses HTAs und würde die Anwendung komplexer Modelle, Annahmen und detaillierte Daten, die für Deutschland nicht vorliegen, erfordern. Um allerdings beantworten zu können, inwieweit dem Teilnahmeverhalten überhaupt eine relevante Rolle bezüglich der Effektivität und Kosten-Effektivität der neuen Screeningtechnologien zukommt, wurde eine einfache Analyse unter Annahme einer zufälligen Teilnahmerate durchgeführt (s. dazu im Methodenteil Abschnitt III C 4. 12).

Zunächst ist festzuhalten, dass die Existenz einer bestimmten Fraktion von Frauen, die grundsätzlich *nie* am Screeningprogramm teilnehmen, die inkrementellen Kosten-Effektivitäts-Verhältnisse nicht verändert, da diese Population keinen Einfluss auf die Inkremente der Berechnungen nimmt. Beeinflusst wird die Kosten-Effektivität der Programme allerdings durch eine mehr oder weniger unregelmäßige Teilnahme am Screeningprogramm. Da für Deutschland keine ausreichenden Daten für die Modellierung bestimmter Screening-Teilnahmeverhalten vorlagen, wurde als einfachstes Szenario eine zufällige Teilnahmerate für jeden Screeningzyklus verwendet. Dies ist bei der Interpretation der im Folgenden dargestellten Ergebnisse zu berücksichtigen (siehe auch Abschnitt III C 4. 12).

In Abbildung 15 werden die diskontierten inkrementellen Kosten-Effektivitäts-Verhältnisse der untersuchten Strategien in Abhängigkeit von den Teilnahmeraten der Frauen am jährlichen Screening für einen Bereich von 0%-100% dargestellt. Es zeigt sich, dass das konventionelle Screeningverfahren im Vergleich zu der Strategie kein Screening bei jeder Teilnahmerate als kosteneffektiv zu bewerten ist. Für die in der Basisfall-Analyse mit 100%-iger Teilnahmerate als nicht kosteneffektiv bewertete computergestützte Auswertung des Papanicolaou-Tests fiel das inkrementelle Kosten-Effektivitäts-Verhältnis bei einer Teilnahmerate von unter 65% unter den Schwellenwert von 50.000 €/LJ. Eine Einführung von Dünnschichtpräparation alleine oder in Kombination mit einer computergestützten Auswertung würde diese Schwelle erst bei einer Teilnahmerate von weniger als circa 40% erreichen, wenn man einen direkten Vergleich mit dem konventionellen Screeningverfahren zugrunde legt.

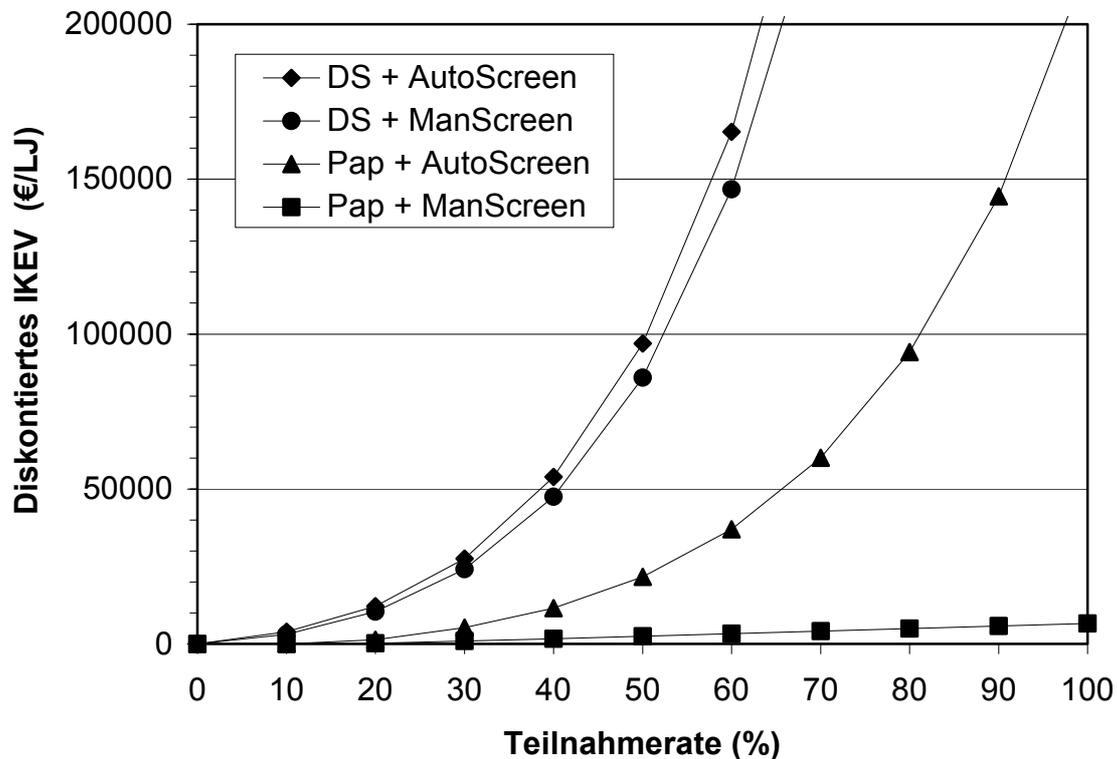


Abbildung 15: Einweg-Sensitivitätsanalyse: Inkrementelle Kosten-Effektivitäts-Verhältnisse (in €/Lebensjahr) der untersuchten Strategien bei unterschiedlichen Teilnahmeraten am Screening.

Pap: Papanicolaou-Test, ManualScreen: manuelle Auswertung, AutoScreen: automatisierte Auswertung, DS: Dünnschichtpräparation

In einer Zwei-Weg-Sensitivitätsanalyse wurden simultan Testkosten und die Werte für die Sensitivität der neuen Screeningverfahren variiert und der Einfluss auf die inkrementellen Kosten-Effektivitäts-Verhältnisse der verschiedenen Strategien untersucht. In Abbildung 16 sind die minimalen und maximalen IKEV der einzelnen Strategien dargestellt. Es zeigte sich für das jährliche Screening, dass auch bei günstigen Werten für die Testkosten und die Sensitivität der neuen Screeningtechnologien keines dieser Verfahren ein akzeptables Kosten-Effektivitäts-Verhältnis aufweist. Selbst unter Annahme der für das IKEV günstigsten Werte der jeweiligen Testkosten und Sensitivität innerhalb der variierten Spannweite liegen die IKEV für alle neuen Screeningtechnologien über 170.000 €/LJ. Im Worst-Case-Szenario dieser Zwei-Weg-Sensitivitätsanalyse wurde die computergestützte Auswertung des Papanicolaou-Tests sogar dominiert, weil sie unter Berücksichtigung der 95%-Konfidenzintervalle für die Sensitivität eine geringere Effektivität als das konventionelle Screeningverfahren mit manueller Auswertung des Papanicolaou-Tests aufwies und dabei aber höhere Kosten verursachte.

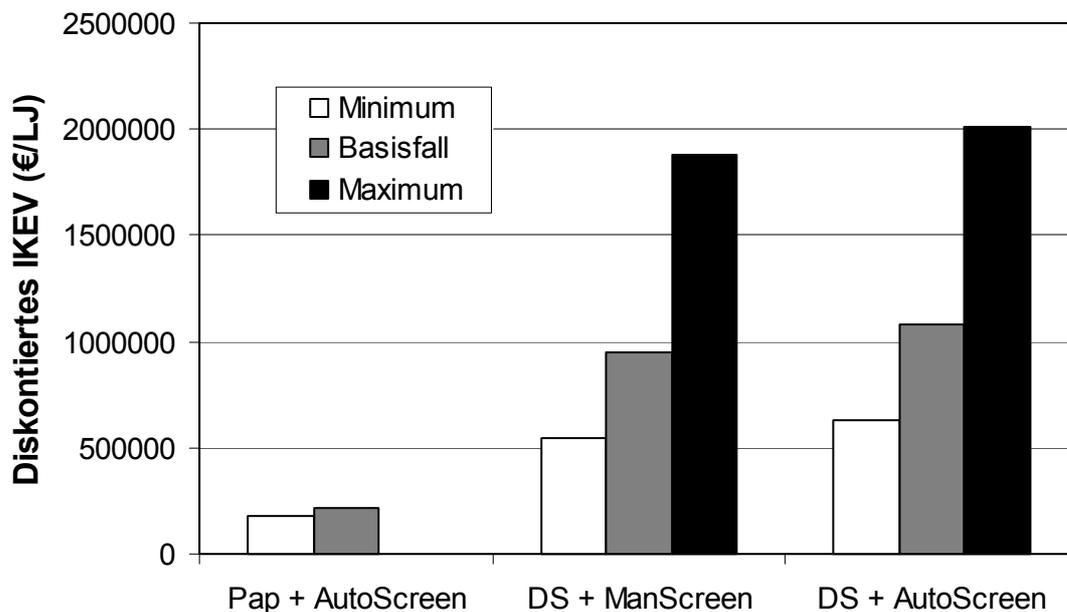


Abbildung 16: Zwei-Weg-Sensitivitätsanalyse: Minimum, Basisfall und Maximum der inkrementellen Kosten-Effektivitäts-Verhältnisse (in €/Lebensjahr) der untersuchten Strategien bei Variation von Kosten und Sensitivitäten bei einjährigem Screeningintervall.

Pap: Papanicolaou-Test, ManualScreen: manuelle Auswertung, AutoScreen: automatisierte Auswertung, DS: Dünnschichtpräparation

Im Fall "Maximum der Strategie Pap + AutoScreen" wurde diese Strategie durch andere Strategien dominiert (d.h. geringere Effektivität bei höheren Kosten); ein IKEV wurde deshalb nicht dargestellt.

III C 6 Diskussion

Für dieses Health Technology Assessment wurde in einem interdisziplinären Netzwerk ein entscheidungsanalytisches Modell zur Evaluation von verschiedenen Screeningprogrammen für das Zervixkarzinom für den Kontext des deutschen Gesundheitssystems entwickelt. Dieses Modell berücksichtigt deutsche epidemiologische Daten aus verschiedenen Krebsregistern, die aktuelle deutsche Praxis in der Diagnose und Behandlung des Zervixkarzinoms und seiner Vorstufen, sowie deutsche Kostendaten. Das Modell konnte für verschiedene Forschungsfragen bezüglich der Evaluation von medizinischen und ökonomischen Langzeit-Konsequenzen eingesetzt werden.

In der Basisfall-Analyse wurde nach aktuellem auf gesetzlicher Basis festgesetztem Krebsfrüherkennungsprogramm ein jährliches Screening ab einem Alter von 20 Jahren festgelegt. Untersucht wurde eine Kohorte von Frauen im Alter von 15 Jahren, für die klinische Ereignisse, Mortalität und Kosten über einen lebenslangen Zeithorizont evaluiert wurden.

In der Basisfall-Analyse der Modellrechnung erhöhte ein Screening mit konventioneller Papanicolaou-Methode die Lebenserwartung um 94 Tage und verhinderte 2.994 Krebsfälle pro 100.000 Frauen. Eine Kombination der Dünnschichtpräparation mit anschließender computerisierter Auswertung erhöhte die Lebenserwartung um einen weiteren halben Tag und verhinderte 3.029 Krebsfälle pro 100.000 Frauen gegenüber keinem Screening. Die anderen Screeningstrategien erzielten Ergebnisse, die zwischen dem konventionellen Verfahren und der Kombination der neuen Verfahren lagen. Das konventionelle Screening erwies sich im Vergleich zu der Strategie kein Screening als kosteneffektiv mit einem akzeptablen Kosten-Effektivitäts-Verhältnis von 6.600 €/LJ. Der Einsatz neuer Screeningverfahren im Vergleich zum konventionellen Verfahren war mit inkrementellen Kosten-Effektivitäts-Verhältnissen zwischen 220.000 €/LJ und 1.083.000 €/LJ bei jährlichem Screening als nicht kosteneffektiv zu bewerten, wenn man dies mit anderen akzeptierten Verfahren in anderen Bereichen der Medizin vergleicht, wie beispielsweise die Hämodialyse (Weinstein 1999) oder die koronare Bypass-Operation (Wong et al. 1990).

III C 6. 1 Screeningintervall

Neben der Basisfall-Analyse mit einjährigem Screeningintervall wurde das Screeningintervall in zusätzlichen Szenario-Analysen auf 2, 3, und 5 Jahre ausgedehnt. Tabelle 60 verdeutlicht, dass mit zunehmender Länge des Screeningintervalls die inkrementelle Lebenserwartung (gewonnene Lebensstage) aller Screeningstrategien im Vergleich zu keinem Screening abnimmt. Das bedeutet, dass das Screening umso effektiver ist, je häufiger man es einsetzt, und dass dieser Effekt bis hin zu einem jährlichen Screeningintervall erhalten bleibt. Allerdings ist der zusätzliche Effekt der neuen Screeningverfahren im Vergleich zum konventionellen Screening am höchsten bei längeren Screeningintervallen. Dies lässt sich folgendermaßen erklären: Bei einem mehrjährigen Screeningintervall mit der konventionellen Papanicolaou-Methode werden aufgrund der relativ schlechten Sensitivität weniger intraepitheliale Läsionen (auch in den höheren Stadien wie CIN 3 oder CIS) entdeckt, die dann zu Zervixkarzinomfällen mit reduzierter Lebenserwartung führen. Deshalb können im Setting mit mehrjährigem Screeningintervall die neuen Screeningtechnologien mit ihrer wenn auch geringfügig erhöhten Sensitivität noch eine nennenswerte Wirkung zeigen, indem durch sie die ansonsten unentdeckten Läsionen entdeckt werden können und schließlich Todesfälle durch Zervixkarzinom verhindert werden können. Der dadurch bedingten Erhöhung der Lebenserwartung stehen zwar erhöhte Gesamtkosten gegenüber, die durch die erhöhten Testkosten und den erhöhten Versorgungsaufwand im Rahmen der Früherkennungs-untersuchungen und deren Nachsorge bedingt sind, jedoch können sich im mehrjährigen Setting noch akzeptable Kosten-Effektivitäts-Verhältnisse für die neuen Screeningtechnologien ergeben. Diese Situation stellt sich anders dar in der deutschen Praxis des einjährigen Screenings: Die in der Routinepraxis hohe Falschnegativ-Rate der konventionellen Papanicolaou-Methode wird dadurch kompensiert, dass dieses Verfahren jedes Jahr angewendet wird. Dadurch reduziert sich die Rate der über mehrere Jahre hinweg bestehenden falsch negativen Testergebnisse und die Gesamtsensitivität dieser Screeningstrategie ist über einige Jahre hinweg so hoch, dass die Zahl der unentdeckten Zervixläsionen oder -krebsfälle bei den regelmäßig am jährlichen Screeningprogramm teilnehmenden Frauen relativ gering ist. Zwar werden nicht unbedingt alle Läsionen im Jahr der Entstehung entdeckt, sondern erst in Folgejahren, dies ist aber in gewissem Rahmen ein durchaus erwünschter Effekt, solange die Läsion überhaupt noch rechtzeitig entdeckt und erfolgreich in einem Frühstadium behandelt werden kann. In diesem Setting können mit den neuen Screeningtechnologien selbst bei einer etwas erhöhten Testsensitivität dieser Verfahren nur wenige zusätzliche Fälle entdeckt werden. Der geringe Gewinn an Lebenserwartung muss allerdings mit erhöhten Testkosten und einem deutlich erhöhten Nachsorgeaufwand bei den höheren jährlichen Entdeckungsraten bezahlt werden, die aber nicht unbedingt zu einer deutlichen Erhöhung der Gesamtsensitivität (über mehrere Jahre hinweg) führen.

Ein wichtiger, im Rahmen dieses HTAs nicht speziell untersuchter unerwünschter Effekt einer zu hohen Sensitivität insbesondere bei den frühen Stadien wie CIN 1 ist eine mögliche Erhöhung bzw. zeitliche Vorziehung von Behandlungen, die nicht immer zwangsläufig mit einer Reduktion der Gesamtmortalität einhergehen muss, und ferner eine Erhöhung der Zeiträume, in denen Frauen wegen vorübergehender positiver Testbefunde in Sorge sind. Da das eingesetzte Modell diese Aspekte nicht explizit berücksichtigt, lassen die vorliegenden Analysen hierzu keine Schlussfolgerungen zu. Diesbezüglich besteht also weiterer Forschungsbedarf, insbesondere ist eine Integration von gesundheitsbezogener Lebensqualität in das entscheidungsanalytische Modell anzustreben.

Den Modellanalysen zufolge bewirkt eine Erhöhung des Screeningintervalls also einerseits eine absolute Reduktion in der Effektivität sowohl des konventionellen Screenings als auch bei den neuen Verfahren, andererseits ergibt sich dadurch eine höhere Effektivität der neuen Screeningverfahren im Vergleich zum konventionellen Verfahren. In der gesundheitsökonomischen Evaluation drückte sich dieses durch geringere Kosten-Effektivitäts-Verhältnisse für die neuen Verfahren im Vergleich mit dem konventionellen Verfahren aus. Bei einem 2-Jahres-Screeningintervall erreichte das Screening mit einer computergestützten Auswertung des Papanicolaou-Tests ein inkrementelles Kosten-Effektivitäts-Verhältnis von ca. 26.000 €/LJ. Bei einem 3-Jahres-Intervall kostete der Einsatz der Dünnschichtpräparation mit manuellem Auswerten 47.000 €/LJ und mit computergestützter Auswertung 54.000 €/LJ. Bei einer gesellschaftlichen Zahlungsbereitschaft von 50.000 €/LJ wäre das computergestützte Auswertungsverfahren bei einem vorgegebenen 2-Jahres-Screeningintervall als kosteneffektiv einzuschätzen, die Einführung der Dünnschichtpräparation mit manueller oder computergestützter Auswertung wäre ab einem vorgegebenen Screeningintervall von 3 Jahren als kosteneffektiv einzuschätzen. Diese Ergebnisse werden durch internationale Studien zur Kosten-Effektivität neuer Screeningverfahren, die im Kapitel II C 5. 4 beschrieben wurden, gestützt (McCroy & Matchar 1999; Payne et al. 2000; Brown & Garber 1999; Hutchinson & Farber 2000).

Weitere wichtige Einblicke ließen sich gewinnen, als eine simultane Betrachtung aller Kombinationen von Screeningintervallen und Screeningtechnologien durchgeführt wurde. Die in diesem Verfahren ermittelte Effizienzlinie (*cost-effectiveness frontier*) unterstützte die Identifikation sogenannter dominierter Strategien, d.h. Strategien, denen eine andere Kombination aus Screeningintervall und Screeningtechnologie sowohl bezüglich der Effektivität als auch bezüglich der Kosten überlegen ist. Aus dieser modellgestützten Analyse konnte abgeleitet werden, dass basierend auf der Evidenz der eingeschlossenen Studien die neuen Technologien im Falle einer erhöhten Testsensitivität selbst im Screening mit 2-Jahresintervall eine ähnliche Effektivität

wie das konventionelle jährliche Screening erbringen könnten und dabei gleichzeitig eine Kosteneinsparung möglich wäre.

Basierend auf diesen Analyseergebnissen ist ein Einsatz der neuen Screeningverfahren anstelle des konventionellen Screenings bei dem aktuell geltenden gesetzlichen Krebsfrüherkennungsprogramms mit jährlichem Screening nicht als kosteneffektiv anzusehen. Es sollte jedoch kritisch in Betracht gezogen werden, ob ein Einsatz der neuen Screeningmethoden in Verbindung mit einer Erhöhung des Screeningintervalls auf 2-Jahre angestrebt werden sollte. Dabei sollten jedoch unbedingt die möglichen Auswirkungen eines solchen Vorgehens auf das Screening-Teilnahmeverhalten berücksichtigt werden.

III C 6. 2 Limitationen und weiterer Forschungsbedarf

Die vorliegende klinische und gesundheitsökonomische Entscheidungsanalyse besitzt verschiedene Limitationen, die ihrerseits einen weiteren Forschungsbedarf bestimmen.

III C 6. 2. 1 Entscheidungsanalytische Modellierung

Im Gegensatz zur Bewertung der intermediären Effektivitäts-Outcomes, die im medizinischen Teil dieses HTA-Reports vorgenommen wurde, wurden in diesem Teil des Berichtes die Ergebnisse einer entscheidungsanalytischen Modellierung beschrieben.

Mittels einer entscheidungsanalytischen Modellierung kann Forschungsfragen nachgegangen werden, für die eine Durchführung von randomisierten klinischen Studien nicht durchführbar ist, einen zu hohen Aufwand darstellt oder zu lange Zeit in Anspruch nehmen würde. Im Falle des Zervixkarzinom-Screenings wäre es extrem aufwendig, ethisch problematisch und in Deutschland schwer durchsetzbar, eine randomisierte klinische Screeningstudie durchzuführen, in der Frauen bezüglich der gestellten Fragestellung mit unterschiedlichen Screeningverfahren und unterschiedlichen Screeningintervallen unter Berücksichtigung von tatsächlichen Ressourcenverbräuchen über die relevante jahrzehntelange Beobachtungszeit hinweg beobachtet werden, um die gewonnene Lebenserwartung bzw. Langzeitmortalität und die eingesetzten Gesamtkosten analysieren zu können. Und selbst wenn eine solche Studie durchgeführt würde, so würden die Studienergebnisse bezüglich der tatsächli-

chen inkrementellen Effektivität durch die unvermeidbaren Wechsel (Noncompliance) der Frauen zwischen den verschiedenen Studienarmen entweder durch Confounding verzerrt oder im besten Falle systematisch unterschätzt werden. Dieses Phänomen wird "Infektion der Screenigarme" genannt.

Bei der Interpretation der beschriebenen Ergebnisse sind jedoch die Limitationen einer solchen Modellierung zu berücksichtigen. Das Modell basiert auf strukturellen Annahmen und numerischen Modellparametern und liefert valide Ergebnisse nur dann, wenn diese Annahmen und Parameter die Wirklichkeit gut widerspiegeln. So basierten die strukturellen Annahmen für die Krankheitsstadieneinteilung im Markov-Modell auf den CIN- und FIGO-Klassifikationen, die eine Vereinfachung eines eher kontinuierlichen Progressionsverlaufes des Zervixkarzinoms darstellen. Ferner lagen für den deutschen Kontext keine detaillierten Progressions- und Regressionsraten für den natürlichen Verlauf des Zervixkarzinoms vor. Diese Parameter mussten deshalb aus internationalen Studien entnommen und an epidemiologische Beobachtungsdaten von deutschen Krebsregistern angepasst werden. Dabei war ein Vorteil, dass verschiedene externe Datenquellen benutzt werden konnten und das Modell einer internen und externen Validierung unterzogen werden konnte, die gute Ergebnisse brachten. Da die Datenanpassung an einer ungescreenten deutschen Population durchgeführt werden musste, wurde auf deutsche Krebsregisterdaten vor Einführung des Zervixkarzinom-Screenings zurückgegriffen. Es ist jedoch möglich, dass diese Daten nicht genau die heutige Situation bezüglich des natürlichen Krankheitsverlaufs widerspiegeln, wie sie heute ohne ein Screening anzutreffen wäre. Insgesamt kann festgestellt werden, dass nach der Validierung des Modells an deutschen Daten und einem Vergleich mit den Ergebnissen internationaler Modelle zum Krankheitsverlauf in westlichen Industrieländern von einer guten Approximation des Modells für den natürlichen Krankheitsverlauf ausgegangen wird.

III C 6. 2. 2 Screening-Teilnahmeverhalten

Die Teilnahmerate am Screening hatte in Sensitivitätsanalysen einen starken Effekt auf das inkrementelle Kosten-Effektivitäts-Verhältnis der neuen Screening-Verfahren im Vergleich zum konventionellen Verfahren. Bei geringeren Teilnahmeraten im jährlichen Screening-Szenario verbesserte sich das Kosten-Effektivitäts-Verhältnis. Bei einer Teilnahmerate von etwa 65% beim jährlichen Screening reduzierte sich das Kosten-Effektivitäts-Verhältnis des Screenings mit computergestützter Auswertung des Papanicolaou-Test und bei einer Teilnahmerate von etwa 40% reduzierte sich das Kosten-Effektivitäts-Verhältnis des Screenings mit Dünnschichtpräparation auf unterhalb des angenommenen Schwellenwertes für die gesellschaftliche Zahlungsbereitschaft von 50.000 €/LJ. Aufgrund fehlender Daten zum Teilnahmeverhalten im

deutschen Kontext wurde im vorliegenden Modell die Teilnahmerate vereinfacht als Teilnahmewahrscheinlichkeit bei jeder Screeninguntersuchung angenommen, die unabhängig von der früheren Screeningteilnahme war (zufällige Teilnahmerate). Es konnte aufgrund fehlender Evidenz also nicht berücksichtigt werden, dass Frauen, die grundsätzlich seltener am Screening teilnehmen eventuell ein höheres Risiko haben, am Screening im darauffolgenden Jahr ebenfalls nicht teilzunehmen (positive Korrelation). Umgekehrt wurde der Effekt nicht berücksichtigt, dass es Frauen gibt, die nach Aussetzen des Screenings in einem Jahr im Folgejahr eine deutlich erhöhte Teilnahmewahrscheinlichkeit besitzen (negative Korrelation). Hier besteht demnach noch Forschungsbedarf einerseits in den Bereichen Psychologie und Public Health, um weitere Evidenz zum Teilnahmeverhalten zu erhalten, und andererseits sind weitere Modellanalysen durchzuführen, in denen der Effekt der verschiedenen Teilnahmeverhaltensweisen analysiert wird. Neben aktuellen Daten zur altersspezifischen Teilnahmerate der Frauen am Zervixscreening sind also auch Fragen wie die folgenden von Interesse: Welche Art des Teilnahmeverhaltens der Frauen am Screening findet sich wie häufig in Deutschland? Gibt es bestimmte Verhaltensmuster und Teilnahmetypen und wie sehen diese aus? Wie reagiert die Teilnahmebereitschaft, wenn das Screeningintervall erhöht wird? Welche Motivationsstrategien könnten die Teilnahmerate am Screening erhöhen? Wie verhält sich die Kosten-Effektivität des Screenings bei einer Steigerung der Teilnahmeraten? Ist ein Programm zur Erhöhung der Teilnahmeraten der Frauen am Screening kosteneffektiv? Ist es aus gesamtgesellschaftlicher Perspektive sinnvoller, Ressourcen in die Erhöhung der Teilnahmeraten bestimmter Subgruppen zu investieren, als diese Ressourcen für die Erhöhung der Sensitivität neuerer, teurerer Screeningverfahren zu verbrauchen?

III C 6. 2. 3 Güterwerte der Screeningtests

Bezüglich der Werte für die Sensitivität und Spezifität der einzelnen Screeningtests besteht sowohl große Unsicherheit als auch vermutlich eine große Heterogenität, was die Qualität der Abstriche und ihre Auswertungen betrifft. Viele der Studien zur Bestimmung der diagnostischen Güte von Screeningtests beziehen entweder Risikopopulationen mit erhöhtem a-priori-Risiko für Zervixkarzinom ein, berichten die Ergebnisse nicht stratifiziert für Krankheitsstadien und unterschiedliche Testschwellenwerte, führen keinen Vergleich mit Goldstandards durch oder unterliegen einem Verification Bias. Für die Ermittlung der entsprechenden Werte bezüglich des konventionellen Papanicolaou-Tests mit manueller Auswertung war es von Vorteil, dass auf ein publiziertes Review zurückgegriffen werden konnte, in dem all diese Probleme berücksichtigt wurden und die Daten disaggregiert und um Verification-Bias bereinigt dargestellt waren, so dass hierzu valide gepoolte Schätzer ermittelt werden konnten. Bezüglich der neuen Testverfahren allerdings gibt es noch wenige qualitativ

hochwertige Studien, die diese Kriterien erfüllen. Im Rahmen dieser Analyse wurden deshalb ausschließlich relative Testgütewerte in Form von gepoolten Odds Ratios aus Vergleichen zwischen konventionellem und den neuen Screeningtests ermittelt und herangezogen, um die Wahrscheinlichkeit für einen Bias zu minimieren. Idealerweise würde man aber auch für die neuen Screeningverfahren große und hochwertige Studien in der Allgemeinbevölkerung für die Besetzung der Modellparameter heranziehen. Diesbezüglich besteht dringender Forschungsbedarf und sobald weitere Daten zu diesem Bereich vorliegen, sollte für die vorliegenden entscheidungsanalytischen Ergebnisse ein Update der Modellierung durchgeführt werden.

III C 6. 2. 4 Screeningtestkosten

Weiterer Forschungsbedarf besteht bezüglich der exakten Kosten der verschiedenen Screeningtests. Die vorliegende Analyse basiert auf den Abrechnungsgebühren für die Untersuchung im Rahmen des gesetzlichen Früherkennungsprogramms und die zytologische Auswertung des Papanicolaou-Tests. Für die neuen Screeningverfahren wurden über einen Microcosting-Ansatz lediglich die zusätzlichen Materialkosten berücksichtigt. Ein potentieller Mehraufwand oder Einsparungen bezüglich der Personalkosten wurden nicht berücksichtigt.

III C 6. 2. 5 Gesundheitsbezogene Lebensqualität

Ein Forschungsbedarf besteht auch in der Erhebung der Lebensqualität von Frauen in den verschiedenen Stadien der Zervixkrankung und der Lebensqualitätseffekt der Durchführung von Konisationen und anderen Behandlungen im Rahmen der Krebsverhütung. Ferner sollte ebenfalls die psychologische Auswirkung positiver Screening-Testbefunde berücksichtigt werden. Lebensqualitätsdaten können in Form von Nutzwerten (Utilities) in entscheidungsanalytische Modelle integriert werden, so dass Kosten-Nutzwert-Analysen durchgeführt werden können. Die Ergebnisse von nach Standard durchgeführten Kosten-Nutzwert-Analysen erlauben eine bessere Vergleichbarkeit der ökonomischen Effizienz von medizinischen Technologien in unterschiedlichen Bereichen des Gesundheitssystems und für unterschiedliche Krankheitsbereiche. Es ist davon auszugehen, dass insbesondere die psychische und physische Beeinträchtigung der Lebensqualität im Rahmen der Vorsorgeuntersuchungen und deren Nachsorge und resultierenden Behandlungen einen nicht zu unterschätzenden Effekt auf die Ergebnisse und deren Interpretation hat, wenn diese einem relativ kleinen Gewinn in der Lebenserwartung gegenübergestellt werden.

III C 6. 2. 6 Die Rolle der Krebsregister

Es ist darauf hinzuweisen, dass diese entscheidungsanalytische Studie ohne die Daten, die den Autoren dieses HTA-Berichts von den Krebsregistern zur Verfügung gestellt wurden, hätte nicht durchgeführt werden können.

Registerdaten wurden in vierfacher Weise in der Modellierung verwendet:

1. Anpassung der Modellparameter an den deutschen epidemiologischen Kontext:

Internationale Daten zur Progression, Regression und Entdeckungsraten wurden anhand deutscher Krebsregisterdaten für eine ungescreente Population an den deutschen epidemiologischen Kontext angepasst. Dadurch war eine Kalibrierung des entscheidungsanalytischen Modells für den deutschen epidemiologischen Kontext möglich.

2. Ermittlung der krankheitsspezifischen Mortalität:

Basierend auf den Daten des Tumorregisters München konnten die jährlichen stadienspezifischen Überlebenswahrscheinlichkeiten für Patientinnen mit einem Zervixkarzinom abgeleitet werden. Diese Zahlen bestimmen in hohem Maße den Benefit eines Screeningverfahrens, da sie direkt in die Berechnungen der inkrementellen Lebenserwartung beim Vergleich verschiedener Screeningverfahren eingehen.

3. Interne Modellvalidierung:

Die standardisierte Berichterstattung der Daten des Gesamtkrebsregisters Berlin ermöglichte eine sorgfältige interne Validierung des entscheidungsanalytischen Modells. Damit konnte gezeigt werden, dass die für die Datenanpassung verwendeten Registerdaten auch tatsächlich präzise durch das Markov-Modell reproduziert wurden.

4. Externe Modellvalidierung:

Von besonders hohem Wert war es, dass zusätzlich zu den für die Kalibrierung verwendeten Daten des Gesamtkrebsregisters Berlin mit den Daten des Krebsre-

gisters Saarland ein zweiter epidemiologischer Datenkörper zur Verfügung stand. Dies ermöglichte die Überprüfung der vom Modell prädizierten Ergebnisse anhand von externen Daten, die nicht für die Modellanpassung verwendet worden sind. Anhand dieses externen Vergleichs zwischen Modellergebnis und tatsächlicher epidemiologischer Evidenz konnte bestätigt werden, dass das im Rahmen dieses HTA-Projektes entwickelte Markov-Modell die Realität valide widerspiegelt.

Diese teils getrennt geführten Register werden auch für weitere epidemiologische Analysen und populationsbasierten Hochrechnungen von hohem Wert sein. Dies gilt ebenso für andere Krebserkrankungen, für die systematische und evidenz-basierte medizinische oder gesundheitsökonomische Health Technology Assessments durchgeführt werden sollen.

III C 7 Schlussfolgerungen

Im Rahmen des vorliegenden HTA konnte ein entscheidungsanalytisches Modell zur Evaluation der klinischen und ökonomischen Langzeit-Konsequenzen von Screeningprogrammen für das Zervixkarzinom für den Kontext des deutschen Gesundheitssystems entwickelt und für verschiedene Forschungsfragen eingesetzt werden.

Im Rahmen der systematischen Entscheidungsanalyse wurden die medizinischen und ökonomischen Konsequenzen des Einsatzes des computergestützten Auswertungsverfahrens, der Dünnschichtpräparation und der Kombination beider Verfahren unter Berücksichtigung der derzeitigen Evidenz und deren Unsicherheit modelliert. Aus den Ergebnissen der Modellanalysen lassen sich die folgenden Schlussfolgerungen ziehen und Empfehlungen ableiten:

- Nach den im Rahmen dieses HTAs durchgeführten Modellanalysen lassen sich im einjährigen Screening durch den Einsatz der computergestützten Auswertung, der Dünnschichtpräparation oder der Kombination beider Verfahren im Vergleich zum konventionellen Screeningverfahren mit der Papanicolaou-Methode nur sehr geringe Gewinne bezüglich der Lebenserwartung, der Verringerung der Krebsfälle und der Reduzierung der krebsbedingten Todesfälle erzielen.
- Im Rahmen der deutschen jährlichen Screeningpraxis ist bei einer regelmäßigen Screeningteilnahme ein Screening mit einer der neuen Technologien nicht als kosteneffektiv zu bewerten. Legt man ein Screeningintervall von zwei oder mehr Jahren zugrunde, so könnte bei einer gesellschaftlichen Zahlungsbereitschaft von 50.000 €/LJ der Einsatz neuer Screeningverfahren eine kosteneffektive Maßnahme darstellen.
- Bei einem zweijährigen Screeningintervall sind die neuen Technologien kaum weniger effektiv als das jährliche konventionelle Screening, aber verursachen zum Teil deutlich weniger Kosten. Unter der Gewährleistung einer ausreichenden Beteiligungsrate der Frauen am Screeningprogramm sollte kritisch in Betracht gezogen werden, ob ein Einsatz der neuen Screeningmethoden in Verbindung mit einer Verlängerung des Screeningintervalls auf 2-Jahre angestrebt werden sollte. Dabei sollten jedoch unbedingt die möglichen Auswirkungen eines solchen Vorgehens auf das Screening-Teilnahmeverhalten berücksichtigt werden.

- Die von den Krebsregistern durchgeführte systematische Erfassung, Dokumentation und Berichterstattung epidemiologischer Daten ist eine notwendige Voraussetzung zur Durchführung von entscheidungsanalytischen populationsbasierten Studien wie dieser. Das Zurückgreifen auf multiple Datenquellen ermöglichte eine externe Modellvalidierung.
- Es besteht dringender Bedarf an qualitativ hochwertigen Studien in repräsentativen Studienpopulationen unter Verwendung eines Goldstandards zur Bestimmung der Sensitivität und Spezifität insbesondere der neuen Screeningverfahren, aber auch des konventionellen Papanicolaou-Verfahrens bei der Routineanwendung im deutschen Kontext.
- Es besteht weiterhin Forschungsbedarf in weiten Bereichen der Erforschung des Teilnahmeverhaltens der Frauen am Screening. Art und mögliche vorliegende Muster des Teilnahmeverhaltens sowie Möglichkeiten einer Erhöhung der Teilnahmebereitschaft und deren medizinische und gesundheitsökonomische Auswirkung sollten untersucht werden. Hierzu kann das im Rahmen dieses HTA entwickelte entscheidungsanalytische Modell eingesetzt werden.

III C 8 Literatur

III C 8. 1 Allgemeine Literatur aus den Kapiteln II und III

1. Abulafia O., Sherer D. M. (1999a). Automated cervical cytology: meta-analyses of the performance of the AutoPap 300 QC System. *Obstet Gynecol Surv* 54 (7), 469-476.
2. Abulafia O., Sherer D. M. (1999b). Automated cervical cytology: meta-analyses of the performance of the PAPNET system. *Obstet Gynecol Surv* 54 (4), 253-64.
3. ACOG (1998). ACOG committee opinion. New Pap test screening techniques. *International Journal of Gynaecology & Obstetrics* 63, 312-4.
4. AHTAC (1998). Review of automated and semi-automated cervical screening devices.: Australian Health Technology Advisory Committee (AHTAC).
5. Albert A. (1981). Estimated cervical cancer disease state incidence and transition rates. *J Natl Cancer Inst* 67 (3), 571-576.
6. Anderl P., Ledermaier O. (1982). [Influence of screening examinations on stage distributions of cervical carcinoma at a gynecological department]. *Wien-Med-Wochenschr* 132 (11), 241-3.
7. Anderson G., Macaulay C., Maticic J., Garner D., Palcic B. (1997). The use of an automated image cytometer for screening and quantitative assessment of cervical lesions in the British Columbia Cervical Smear Screening Programme. *Cytopathology* 8 (5), 298-312.
8. Anderson T. L., Nelson A. C. (1994). A rational approach to proficiency testing. *Laboratory Medicine* 25 (4), 274-276.
9. Anonymous (1992). The Revised Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses: Report of the 1991 Bethesda Workshop. *The Journal of Reproductive Medicine* 37 (5), 383-386.
10. Anonymous (1996). Improving cervical screening with PAPNET. *European Journal of Cancer Care* 5 (1), 3.
11. Anonymous (1998). Semi-automatic system for cancer screening cytology. *Gynäkologe* 31 (2), 193-195.

12. Anonymous (1999a). Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation - Revidierte Fassung des Hannoveraner Konsensus. Hannoveraner Konsensus Gruppe. *Gesundheitsökonomie und Qualitätsmanagement* 4, A62-A65.
13. Anonymous (1999b). Empfehlungen zur Diagnostik, Therapie und Nachsorge: Zervixkarzinom.: Tumorzentrum München an den Medizinischen Fakultäten der Ludwig-Maximilians-Universität und der Technischen Universität, pp. 98.
14. Anonymous (2002). Frauengesundheitsbericht Bremen 2001. In Der Senator für Arbeit, Frauen, Gesundheit, Jugend und Soziales. Abteilung Gesundheitswesen Referat Gesundheitsberichterstattung. <http://www.bremen.de/info/frauengesundheitsbericht/>, Freie Hansestadt Bremen, pp. 57-65.
15. Aponte-Cipriani S. L., Teplitz C., Rorat E., Savino A., Jacobs A. J. (1995). Cervical smears prepared by an automated device versus the conventional method. A comparative analysis. *Acta Cytol* 39 (4), 623-30.
16. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) (2000). Kurzgefasste Interdisziplinäre Leitlinien 2000: Zervixkarzinom. <http://www.uni-duesseldorf.de/WWW/AWMF/ziplinäre>, 269ff.
17. Arends M. J., Buckley C. H., Wells M. (1998). Aetiology, pathogenesis, and pathology of cervical neoplasia. *J Clin Pathol* 51 (2), 96-103.
18. Ashfaq R., Gibbons D., Vela C., Saboorian M. H., Iliya F. (1999). ThinPrep Pap Test. Accuracy for glandular disease. *Acta Cytol* 43 (1), 81-85.
19. Ashfaq R., Liang Y., Saboorian M. H. (1995). Evaluation of PAPNET system for rescreening of negative cervical smears. *Diagn Cytopathol* 13 (1), 31-36.
20. Ashfaq R., Saliger F. (1993). Evaluation of the PAPNET system for prescreening triage. *Acta Cytol* 37, 771.
21. Ashfaq R., Saliger F., Solares B., Thomas S., Liu G., Liang Y., Saboorian M. H. (1997). Evaluation of the PAPNET system for prescreening triage of cervicovaginal smears. *Acta Cytol* 41 (4), 1058-1064.
22. Austin R. M. (1998). Implementing liquid-based gynecologic cytology: balancing marketing, financial, and scientific issues. *Cancer* 84 (4), 193-196.
23. Austin R. M., Ramzy I. (1998). Increased detection of epithelial cell abnormalities by liquid-based gynecologic cytology preparations. A review of accumulated data. *Acta Cytol* 42 (1), 178-184.
24. Awen C., Hathway S., Eddy W., Voskuil R., Janes C. (1994). Efficacy of Thin-Prep preparation of cervical smears: a 1,000-case, investigator-sponsored study. *Diagn Cytopathol* 11 (1), 33-36.

25. Bai H., Sung C. J., Steinhoff M. M. (2000). ThinPrep Pap Test promotes detection of glandular lesions of the endocervix. *DiagnCytopathol* 23 (1), 19-22.
26. Baird P. J. (1997). Evaluation of the PAPNET system in a general pathology service. *Med J Aust* 167 (5 UR - PM:9315021), 285-286.
27. Bajardi F. (1979). [Combined use of colposcopy and cytology]. *Fortschr-Med* 97 (38), 1691-5.
28. Baldauf J. J., Dreyfus M., Lehmann M., Ritter J., Philippe E. (1995). Cervical cancer screening with cervicography and cytology. *European Journal of Obstetrics, Gynecology, & Reproductive Biology* 58 (1), 33-9.
29. Baltzer J. (1998). Diagnostic methods and therapy in pre- and early stages of cervical carcinoma. *Onkologie* 4 (2), 124-134.
30. Bappert L. (1982). [Efficacy of cancer screening examination in women]. *MMW-Munch-Med-Wochenschr* 124 (50), 1124-8.
31. Bartels P. (1999). Automation of primary screening for cervical cancer. sooner or later? *Acta Cytol* 43, 7-12.
32. Bartels P. H. (2000). Automated primary screening devices. Expectations for the next generation. *Acta Cytol* 44 (5), 703-8.
33. Bartels P. H., Bibbo M., Hutchinson M. L., Gahm T., Grohs H. K., Gwi-Mak E., Kaufman E. A., Kaufman R. H., Knight B. K., Koss L. G., et al. (1998). Computerized screening devices and performance assessment: development of a policy towards automation. International Academy of Cytology Task Force summary. *Diagnostic Cytology Towards the 21st Century: An International Expert Conference and Tutorial. Acta Cytol* 42 (1), 59-68.
34. Bartoo G. T., Lee J. S. J., Bartels P. H., Kiviat N. B., Nelson A. C. (1992). Automated prescreening of conventionally prepared cervical smears: A feasibility study. *Laboratory Investigation* 66 (1), 116-122.
35. Bayrle W. (1977). [Critical observations on the rate of "False Negative" findings in gynecologic cytology (author's transl)]. *Geburtshilfe-Frauenheilkd* 37 (10), 864-8.
36. Beck J. R., Pauker S. G. (1983). The Markov process in medical prognosis. *Med Decis Making* 3 (4), 419-458.
37. Beckmann M. W., Werner Y., Renner S. P., Fasching P. A., Jap D., Kuschel B. (2000). Cancer detection in gynecology. *Gynäkologe* 33 (7), 474-482.
38. Bednov Y. M., Brosky K. R., Sidawy M. K. (1998). Comparison of the efficacy of the ThinPrep method with the conventional cervical smear. *Laboratory Investigation* 78 (1), 191.

39. Bedrossian C. (1998). Proposed guidelines for secondary screening (rescreening) instruments for gynaecologic cytology. Intersociety Working Group for Cytology Technologies. *Acta Cytol* 42, 1311-4.
40. Beier H. J., Kranzfelder D. (1989). The value of endometrial cytology in screening of gynecological cancers. *Archives of Gynecology and Obstetrics* 245, 1-4.
41. Bell S. (1995). Psychological Response to cervical cancer screening. *Preventive Medicine* 24, 610-6.
42. Bergeron C., Masseroli M., Ghezi A., Lemarie A., Mango L., Koss L. G. (2000). Quality control of cervical cytology in high-risk women. PAPNET system compared with manual rescreening. *Acta Cytol* 44 (2), 151-157.
43. Berndt H., Hendel V. (1984). [The incidence of malignant neoplasms of the female genital organs in East Germany]. *Zentralbl Gynakol* 106 (3), 145-51.
44. Bibbo M., Hawthorne C., Zimmerman B. (1999). Does use of the AutoPap assisted primary screener improve cytologic diagnosis? *Acta Cytol* 43 (1), 23-26.
45. Bierfreund B. (1986). [Results of the study of 96,483 cervical smears of young females up to 30 years of age. A review 5 years following the expansion of a preventive care program]. *Geburtshilfe-Frauenheilkd* 46 (11), 804-7.
46. Bishop J. W. (1997). Comparison of the CytoRich system with conventional cervical cytology. Preliminary data on 2,032 cases from a clinical trial site. *Acta Cytol* 41 (1), 15-23.
47. Bishop J. W., Bigner S. H., Colgan T. J., Husain M., Howell L. P., McIntosh K. M., Taylor D. A., Sadeghi M. H. (1998). Multicenter masked evaluation of AutoCyte PREP thin layers with matched conventional smears. Including initial biopsy results. *Acta Cytol* 42 (1), 189-197.
48. Bishop J. W., Chevront D. A., Elston R. J. (1999a). Utility of residual AutoCyte cervical cytology samples for image analysis. *Acta Cytol* 43 (1), 39-46.
49. Bishop J. W., Chevront D. A., Sims K. L. (2000). Evaluation of the AutoCyte SCREEN system in a clinical cytopathology laboratory. *Acta Cytol* 44 (2), 128-136.
50. Bishop J. W., Kaufman R. H., Taylor D. A. (1999b). Multicenter comparison of manual and automated screening of AutoCyte gynecologic preparations. *Acta Cytol* 43 (1), 34-38.
51. Blohmer J. U. (1999). Screening and therapy of early cervical carcinoma. *Onkologie* 2, 25-26.

52. Bolick D. R., Hellman D. J. (1998). Laboratory implementation and efficacy assessment of the ThinPrep cervical cancer screening system. *Acta Cytol* 42 (1), 209-213.
53. Boon M. E., Kok L. P. (1993). Neural network processing can provide means to catch errors that slip through human screening of Pap smears. *Diagn Cytopathol* 9, 411-16.
54. Boon M. E., Kok L. P., Beck S. (1995). Histologic validation of neural network-assisted cervical screening: comparison with the conventional procedure. *Cell Vision* 2, 23-7.
55. Boon M. E., Kok L. P., Nygaard-Nielsen M., Holm K., Holund B. (1994). Neural network processing of cervical smears can lead to a decrease in diagnostic variability and an increase in screening efficacy: a study of 63 false-negative smears. *Mod Pathol* 7 (9), 957-961.
56. Bosanquet N., Coleman D. V., Dore C. J., Douglas G., Magee L. J., Baker R., Amerasinghe C., Falcon C., McKee G., Sayer B., et al. (1999). Assessment of automated primary screening on PAPNET of cervical smears in the PRISMATIC trial. *Lancet* 353 (9162), 1381-1385.
57. Braun K. (1999). New technologies in cervical screening: Are they necessary? (<http://www.womhealthorgau/healthjourney/papsmeartechnologyhtm>).
58. Brenner H., Siegle S., Stegmaier C., Ziegler H. (1993). Second primary malignancies following gynecological tumors in Saarland, Germany, 1968-1987. *Journal of Cancer Research & Clinical Oncology* 119 (3), 179-183.
59. Brenner H., Stegmaier C., Ziegler H. (1990a). [Incidence of secondary tumors following cervical neoplasia in Saarland 1968-1987]. *Geburtshilfe und Frauenheilkunde* 50 (8), 614-8.
60. Brenner H., Wiebelt H., Ziegler H. (1990b). [Progress in the early detection of cervix cancer from the viewpoint of the Saarland cancer register]. *Geburtshilfe und Frauenheilkunde* 50 (4), 304-9.
61. Broadstock M. (2000). Effectiveness and cost-effectiveness of automated and semi-automated cervical screening devices: a systematic review. In <http://nzhta.chmeds.ac.nz/csv3n1.htm>: New Zealand Health Technology Assessment (NZHTA).
62. Brotzman G. L., Kretzchmar S., Ferguson D., Gottlieb M., Stowe C. (1999). Costs and outcomes of PAPNET secondary screening technology for cervical cytologic evaluation. A community hospital's experience. *Arch Fam Med* 8 (1), 52-55.

63. Brown A. D., Garber A. M. (1998). Cost-effectiveness of three new technologies to enhance Pap testing. Chicago: Blue Cross & Blue Shield Association Technology Evaluation Center.
64. Brown A. D., Garber A. M. (1999). Cost-effectiveness of 3 methods to enhance the sensitivity of Papanicolaou testing. *Jama* 281 (4), 347-353.
65. Bur M., Knowles K., Pekow P., Corral O., Donovan J. (1995). Comparison of ThinPrep preparations with conventional cervicovaginal smears. Practical considerations. *Acta Cytol* 39 (4), 631-642.
66. Burghardt E., Östör A. G. (1983). Site And Origin Of Squamous Cervical Cancer: A Histomorphologic Study. *Obstetrics & Gynaecology* 62 (1), 117-127.
67. Burghardt E., Winter R., Tamussino K., Pickel H., Lahousen M., Haas J., Girardi F., Ebner F., Hackl A., Pfister H. (1994). Diagnosis and surgical treatment of cervical cancer. *Crit Rev Oncol Hematol* 17, 181-231.
68. Canavan T. P., Doshi N. R. (2000). Cervical cancer. *Am Fam Physician* 61 (5), 1369-1376.
69. Carpenter A. B., Davey D. D. (1999). ThinPrep Pap Test: performance and biopsy follow-up in a university hospital. *Cancer* 87 (3), 105-112.
70. Carter P. M., Coburn T. C., Luszczak M. (1993). Cost-effectiveness of cervical cytologic examination during pregnancy. *J Am Board Fam Pract* 6 (6), 537-545.
71. Carter S. L. (1994). Screening tests for cervical abnormalities. *Nurse Practitioner* 19 (6), 28-2.
72. Cenci M., Nagar C., Giovagnoli M. R., Vecchione A. (1997). The PAPNET system for quality control of cervical smears: validation and limits. *Anticancer Res* 17 (6D), 4731-4734.
73. Cenci M., Nagar C., Vecchione A. (2000). PAPNET-assisted primary screening of conventional cervical smears. *Anticancer Res* 20 (5C), 3887-3889.
74. Cenci M., Vecchione A. (2000). Usefulness of cervical collection by the exact touch, the saccomanno single sampler, combined with automated primary screening. *Diagn Cytopathol* 23 (4), 242-244.
75. Check W. (1997). Too early to solve Pap device puzzle. *CAP Today* 11 (6), 1, 44-1, 49.
76. Chevront D. A., Elston R. J., Bishop J. W. (1998). Cytology. Effect of a thin-layer preparation system on workload in a cytology laboratory. *Laboratory Medicine* 29 (3), 174-179.
77. Coleman D. (1998). Evaluation of automated systems for the analysis of cervical smears. *Cytopathology* 9, 359-368.

78. Colgan T. J., Bon N., Lee J. S., Patten S. F. (1997). AutoPap 300 QC system scoring of cervical smears without "epithelial cell abnormalities". *Acta Cytol* 41 (1), 45-49.
79. Colgan T. J., Patten S. F., Lee J. S. (1995). A clinical trial of the AutoPap 300 QC system for quality control of cervicovaginal cytology in the clinical laboratory. *Acta Cytol* 39 (6), 1191-1198.
80. Corkill M., Knapp D., Martin J., Hutchinson M. L. (1997). Specimen adequacy of ThinPrep sample preparations in a direct-to-vial study. *Acta Cytol* 41 (1), 39-44.
81. Cuzick J., Sasieni P., Davies P., Adams J., Normand C., Frater A., van Ballegooijen M., van den Akker E. (1999). A systematic review of the role of human papillomavirus testing within a cervical screening programme. In (<http://www.ncchta.org/htapubs.htm>): The National Coordinating Centre for Health Technology Assessment (NCCHTA), pp. 204.
82. Cuzick J., Sasieni P., Davies P., Adams J., Normand C., Frater A., van Ballegooijen M., van den Akker-van Marle M. E. (2000). A systematic review of the role of human papilloma virus (HPV) testing within a cervical screening programme: summary and conclusions. *Br J Cancer* 83 (5), 561-565.
83. Diaz-Rosario L. A., Kabawat S. E. (1999). Performance of a fluid-based, thin-layer papanicolaou smear method in the clinical setting of an independent laboratory and an outpatient screening population in New England. *Arch Pathol Lab Med* 123 (9), 817-821.
84. DiBonito L., Falconieri G., Tomasic G., Colautti I., Bonifacio D., Dudine S. (1993). Cervical cytopathology: An evaluation of its accuracy based on cytohistologic comparison. *Cancer* 72 (10), 3002-3006.
85. Dietl J., Buchholz F., Semm K. (1981). [Epidemiological and diagnostic analyses of 1194 cone biopsies]. *Geburtshilfe-Frauenheilkd* 41 (3), 173-178.
86. Doornewaard H., van de Seijp H., Woudt J. M., van der Graaf Y., van den Tweel J. G. (1997a). Negative cervical smears before CIN 3/carcinoma. Reevaluation with the PAPNET Testing System. *Acta Cytol* 41 (1), 74-78.
87. Doornewaard H., van der Schouw Y. T., van der Graaf Y. (2000). Reproducibility in double scanning of cervical smears with the PAPNET system. *Acta Cytol* 44 (4), 604-610.
88. Doornewaard H., van der Schouw Y. T., van der Graaf Y., Bos A. B., Habbema J. D., van den Tweel J. G. (1999). The diagnostic value of computer-assisted primary cervical smear screening: a longitudinal cohort study. *Mod Pathol* 12 (11), 995-1000.

89. Doornewaard H., Woudt J. M., Strubbe P., van de Seijp H., van den Tweel J. G. (1997b). Evaluation of PAPNET-assisted cervical rescreening. *Cytopathology* 8 (5), 313-321.
90. Drummond M. F., O'Brien B., Stoddart G. L., Torrance G. W. (1997). *Methods for the Economic Evaluation of Health Care Programmes*, 2nd edn. Oxford University Press Inc., New York. pp. 305.
91. Duggan M. A. (2000). Papnet-assisted, primary screening of cervico-vaginal smears. *Eur J Gynaecol Oncol* 21 (1), 35-42.
92. Duggan M. A., Brasher P. (1997). Paired comparison of manual and automated Pap test screening using the PAPNET system. *Diagn Cytopathol* 17 (4), 248-254.
93. Dunton C. J. (2000). New technology in Papanicolaou smear processing. *Clinical Obstetrics and Gynecology* 43 (2), 410-417.
94. Dupree W. B., Suprun H. Z., Beckwith D. G., Shane J. J., Lucente V. (1998). The promise and risk of a new technology: The Lehigh Valley Hospital's experience with liquid-based cervical cytology. *Cancer* 84 (4), 202-207.
95. Ebeling K. (1983). [Screening in gynecologic oncology]. *Zentralbl Gynakol* 105 (19), 1225-1238.
96. Ebeling K. (1986). [Practical points of reference for the prevention and early detection of cervical cancer]. *Z Arztl Fortbild Jena* 80 (7), 306-310.
97. Ebeling K. (1987). [Early diagnosis of cervical cancer]. *Z Arztl Fortbild Jena* 81 (15), 763-769.
98. Ebeling K., Nischan P. (1987). [Epidemiology of cervix cancer--a review]. *Arch Geschwulstforsch* 57 (5), 401-416.
99. Ebeling K., Sarembe B. (1986). [Principles of the prevention, early detection, treatment and aftercare of cervix cancer]. *Zentralbl Gynakol* 108 (23), 1401-1411.
100. Ebeling K., Weissbach A., Nischan P. (1987). [Microinvasive stage Ia cancer of the uterine cervix--a population based analysis]. *Zentralbl Gynakol* 109 (10), 634-646.
101. ECRI (1998a). Automated monolayer slide preparation devices for Pap smear screening. Plymouth Meeting, PA: ECRI Health Technology Assessment Information Service,.
102. ECRI (1998b). Automated Pap smear rescreening technologies Plymouth Meeting, PA: ECRI Health Technology Assessment Information Service,.

103. ECRI (1999). Automated monolayer slide preparation system for pap smear screening: ThinPrep 2000. Plymouth Meeting, PA: ECRI Health Technology Assessment Information Service,.
104. ECRI (2000). Automated pap smear screening technologies: the autopap system. Plymouth Meeting, PA: ECRI Health Technology Assessment Information Service,.
105. Eddy D. M. (1987). The frequency of cervical cancer screening. Comparison of a mathematical model with empirical data. *Cancer* 60 (5), 1117-1122.
106. Eddy D. M. (1990). Screening for cervical cancer. *Ann Intern Med* 113 (3), 214-226.
107. Fahey M. T., Irwig L., Macaskill P. (1995). Meta-Analysis of Pap test accuracy. *Am J Epidemiol* 141 (7), 680-689.
108. Fahs M. C., Mandelblatt J., Schechter C., Muller C. (1992). Cost effectiveness of cervical cancer screening for the elderly. *Ann Intern Med* 117 (6), 520-527.
109. Farnsworth A., Chambers F. M., Goldschmidt C. S. (1996). Evaluation of the PAPNET system in a general pathology service. *Med J Aust* 165 (8), 429-431.
110. Ferenczy A., Robitaille J., Franco E., Arseneau J., Richart R. M., Wright T. C. (1996). Conventional cervical cytologic smears vs. ThinPrep smears. A paired comparison study on cervical cytology. *Acta Cytol* 40 (6), 1136-1142.
111. Ferris D. G., Heidemann N. L., Litaker M. S., Crosby J. H., Macfee M. S. (2000). The efficacy of liquid-based cervical cytology using direct-to- vial sample collection. *Journal of Family Practice* 49 (11), 1005-1011.
112. Fetterman B., Pawlick G., Koo H., Hartinger J., Gilbert C., Connell S. (1999). Determining the utility and effectiveness of the NeoPath AutoPap 300 QC System used routinely. *Acta Cytol* 43 (1), 13-22.
113. Foreman J., Hader J. (1997). Cervical cancer screening. In <http://www.sdh.sk.ca/hsurc/cervical.htm>: Health Services Utilization and Research Committee (HSURC).
114. Gatta G., Capocaccia R., Hakulinen T., Sant M., Verdecchia A., De Angelis G., Micheli A., Berrino F. (1999). Variations in survival for invasive cervical cancer among European women, 1978-89. EURO CARE Working Group. *Cancer Causes Control* 10 (6), 575-581.
115. Gay D. J., Donaldson L. D., Goellner J. R. (1985). False-Negative Results in Cervical Cytologic Studies. *Acta Cytol* 29 (6), 1043-1046.
116. Geyer J. W., Hancock F., Carrico C., Kirkpatrick M. (1993). Preliminary evaluation of Cyto-Rich: an improved automated cytology preparation. *DiagnCytopathol* 9 (4), 417-422.

117. Ghidoni D., Fabbris E., Folicaldi S., Amadori A., Medri M., Bucchi L., Bondi A. (1998). Accuracy comparison between PAPNET diagnoses and conventional diagnoses in an Italian cervical cytology laboratory. *Diagn Cytopathol* 19 (4), 279-283.
118. Gilman J., Myatt M. (1998). *EpiCalc 2000* (computer program): Brixton Books.
119. Gitsch G., Kainz C., Breitenecker G. (1993). [The incidence of positive cervix cytology 1980-1989]. *Gynakol Geburtshilfliche Rundsch* 1 (221), 1018-8843.
120. Gold M. R., Siegel J. E., Russell L. B., Weinstein M. C. (1996). *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York.
121. Goldie S., Kuhn L., Denny L., Pollack A., Wright T. C. (2001). Policy Analysis of Cervical Cancer Screening Strategies in Low-Resource Settings: Clinical Benefits and Cost-effectiveness. *JAMA* 285 (24), 3107-3115.
122. Goldie S. J., Weinstein M. C., Kuntz K. M., Freedberg K. A. (1999). The costs, clinical benefits, and cost-effectiveness of screening for cervical cancer in HIV-infected women. *Ann Intern Med* 130 (2), 97-107.
123. Gornall R. J., Singh N., Noble W., Boyd I. E., Hitchcock A. (2000). Glandular abnormalities on cervical smear: a study to compare the accuracy of cytological diagnosis with underlying pathology. *European Journal of Gynaecological Oncology* 21 (1), 49-52.
124. Greiner W., Hoffmann C. (1999). Leitlinien zur gesundheitsökonomischen Evaluation. In *Ansätze und Methoden der ökonomischen Evaluation - eine internationale Perspektive*. Health Technology Assessment, Leidl R, von der Schulenburg, J.M., Wasem, J., ed. Nomos, Baden-Baden, pp. 129-155.
125. Greiner W., Uber A. (2000). Gesundheitsökonomische Studien und der Einsatz von Lebensqualitätsindices. In *Lebensqualitätsforschung und Gesundheitsökonomie in der Medizin*, Ravens-Sieberer U, Cieza A, von Steinbüchel N, Bullinger M, eds. Ecomed, Landsberg.
126. Grussendorf Conen E. I. (1993). [Human pathogenic papillomaviruses in development of cervix cancer. Practical consequences in preventive screening of women]. *Fortschr-Med* 111 (9), 145-146.
127. Guidos B. J., Selvaggi S. M. (1999). Use of the Thin Prep (R) Pap Test (TM) in clinical practice. *Diagn Cytopathol* 20 (2), 70-73.
128. Guidos B. J., Selvaggi S. M. (2000). Detection of endometrial adenocarcinoma with the ThinPrep(R) pap testtrade mark. *Diagn Cytopathol* 23 (4), 260-265.
129. Gullotta G., Margariti P. A., Rabitti C., Balsamo G., Valle D., Capelli A., Mancuso S. (1997). Cytology, histology, and colposcopy in the diagnosis of neoplastic non-invasive epithelial lesions of the cervix. *European Journal of Gynaecological Oncology* 18 (1), 36-38.

130. Gustafsson L., Ponten J., Bergstrom R., Adami H. O. (1997). International incidence rates of invasive cervical cancer before cytological screening. *International Journal of Cancer* 71 (2), 159-165.
131. Halford (1998). Cervical Cytology- a review of evolving technologies. *Australian Journal of Medical Science* 19, 8-19.
132. Halford J. A., Wright R. G., Ditchmen E. J. (1997). Quality assurance in cervical cytology screening. Comparison of rapid rescreening and the PAPNET Testing System. *Acta Cytol* 41 (1), 79-81.
133. Halford J. A., Wright R. G., Ditchmen E. J. (1999). Prospective study of PAPNET: review of 25,656 Pap smears negative on manual screening and rapid rescreening. *Cytopathology* 10 (5), 317-323.
134. Heinzl S. (1989). [Cervical intra-epithelial neoplasia]. *Ther Umsch* 46, 839-846.
135. Heinzl S., Hirsbrunner A. G. (1991). Screening patterns among women attending the University Gynaecology Clinic in Basel. *Schweizerische Medizinische Wochenschrift* 121 (11), 378-384.
136. Hessel F., Kohlmann T., Krauth C., Novy R., Seitz R., Siebert U., Wasem J. (1999). Gesundheitsökonomische Evaluation in der Rehabilitation. In: Verband Deutscher Rentenversicherungsträger (Hrsg.): *Förderschwerpunkt "Rehabilitationswissenschaften". Empfehlungen der Arbeitsgruppen "Generische Methoden", "Routinedaten" und "Reha-Ökonomie"*. DRV-Schriften Bd. 16, 106-246.
137. Hildesheim A., Schiffman M., Gravitt P. (1994). Persistence of type-specific human papillomavirus infection among cytologically normal women. *J Infect Dis* 169, 235-240.
138. Hillemanns P., Kimmig R., Dannecker C., Noorzai T., Diebold J., Thaler C. J., Hepp H. (2000). LEEP versus cold knife conization in the therapy of cervical intraepithelial neoplasia. *Zentralblatt für Gynäkologie* 122 (1), 35-42.
139. Hillemanns P., Thaler C., Kimmig R. (1997). [Epidemiology and diagnosis of cervical intraepithelial neoplasia--is the present concept of screening and diagnosis still current?]. *Gynäkol Geburtshilfliche Rundsch* 37 (4), 179-190.
140. Ho G. Y. F., Bierman R., Beardsley L. (1998). Natural history of cervicovaginal papillomavirus infection in young women. *N Engl J Med* 338, 423-428.
141. Hochuli E., Rageth C. (1990). Efficiency of gynaecological screening - Perspectives. *Geburtshilfe und Frauenheilkunde* 50 (6), 483-487.
142. Hoerl H. D., Shalkham J. E., Cheung K., Hurlbert S. D., Inhorn S. L., Kurtycz D. F. (2000). Screening parameters for ThinPrep and conventional gynecologic cytology via automated monitoring. *Acta Cytol* 44 (4), 618-624.

143. Hofmann R., Barten G., Rohde E., Beust M., Hamann H. (1989). Frequency, causes and results of suspect vaginal smears. *Zentralblatt für Gynäkologie* 111 (23), 1549-1554.
144. Horn L. C., Kuhndel K., Bilek K. (1995). Histopathologic correlation of false positive cytologic findings in the cervix. *Geburtshilfe und Frauenheilkunde* 55 (8), 425-430.
145. Howell L. P., Belk T., Agdigos R., Davis R., Lowe J. (1999). AutoCyte Interactive Screening System. Experience at a university hospital cytology laboratory. *Acta Cytol* 43 (1), 58-64.
146. Howell L. P., Davis R. L., Belk T. I., Agdigos R., Lowe J. (1998). The AutoCyte preparation system for gynecologic cytology. *Acta Cytol* 42 (1), 171-177.
147. HTAC (1999). *New Technologies for Cervical Cancer Screening.*: Health Technology Advisory Committee (HTAC).
148. Huang T. W., Lin T. S., Lee J. S. (1999). Sensitivity studies of AutoPap System Location-Guided Screening of cervical-vaginal cytologic smears. *Acta Cytol* 43 (3), 363-368.
149. Hutchinson M. L. (1996a). Assessing costs and benefits of PAPNET - Reply. *Acta Cytol* 40 (5), 1109-1111.
150. Hutchinson M. L. (1996b). Assessing the costs and benefits of alternative re-screening strategies. *Acta Cytol* 40, 4-8.
151. Hutchinson M. L., Agarwal P., Denault T., Berger B., Cibas E. S. (1992). A new look at cervical cytology. ThinPrep multicenter trial results. *Acta Cytol* 36 (4), 499-504.
152. Hutchinson M. L., Berger B. M., Farber F. L. (2000). Clinical and cost implications of new technologies for cervical cancer screening: the impact of test sensitivity. *Am J Manag Care* 6 (7), 766-780.
153. Hutchinson M. L., Cassin C. M., Ball H. G. (1991). The efficacy of an automated preparation device for cervical cytology. *Am J Clin Pathol* 96 (3), 300-305.
154. Hutchinson M. L., Farber F. L. (2000). Cost-effectiveness of new cytology technologies. *Am J Clin Pathol* 113 (1), 149-51.
155. Hutchinson M. L., Wilbur D. C., Lapidus S. N. (1995). BLINDED COMPARISON OF CERVICAL CYTOLOGY PREPARATION TECHNIQUES - CLINICAL-TRIALS OF THINPREP(R) VS SMEARED SLIDES. *Laboratory Investigation* 72 (1), 220.

156. Hutchinson M. L., Zahniser D. J., Sherman M. E., Herrero R., Alfaro M., Bratti M. C., Hildesheim A., Lorincz A. T., Greenberg M. D., Morales J., Schiffman M. (1999). Utility of liquid-based cytology for cervical carcinoma screening: results of a population-based study conducted in a region of Costa Rica with a high incidence of cervical carcinoma. *Cancer* 87 (2), 48-55.
157. Hyde C., Cuzick J., Sasieni P. (1999). Cervical screening programs: can automation help? Evidence from a systematic review and economic modelling, applied to the UK - secondary research (project). In <http://www.hta.nhsweb.nhs.uk/>: UK NHS National Coordinating Centre for Health Technology Assessment (NCCHTA).
158. Inhorn S. L. (1998). Validation of the ThinPrep Papanicolaou test for cervical cancer diagnosis. *J lower Genital Tract Disease* 2, 208-12.
159. James L., Knapp D. (1992). New cervical cytology preparation device shows improved sensitivity to abnormal findings over the conventional papanicolaou smear. *Acta Cytol* 36, 579.
160. Jenny J., Isenegger I., Boon M. E., Husain O. A. (1997). Consistency of a double PAPNET scan of cervical smears. *Acta Cytol* 41 (1), 82-87.
161. Johnson N., Sutton J., Thornton J. G., Lilford R. J., Johnson V. A., Peel K. R. (1993). Decision analysis for best management of mildly dyskaryotic smear. *Lancet* 342 (8863), 91-96.
162. Jones B. A., Novis D. A. (1996). Cervical biopsy-cytology correlation: A College of American Pathologists Q-Probes study of 22 439 correlations in 348 laboratories. *Archives of Pathology & Laboratory Medicine* 120 (6), 523-531.
163. Kainz C., Tempfer C., Bancher D., Sliutz G., Breitenecker G., Reinthaller A. (1995). [The Bethesda system--an improvement in classification of cervix cytology?]. *Geburtshilfe und Frauenheilkunde* 55 (8), 435-40.
164. Kaminsky F. C., Burke R. J., Haberle K. R., Mullins D. L. (1995). Rescreening policies in cervical cytology and their effect on detecting the truly positive patient. *Acta Cytol* 39 (2), 239-245.
165. Kaser O., Haller U., Heinzl S., Szalmay G. (1981). [Diagnostic and therapeutic procedures in cases of pathological Papanicolaou smears and suspicion of severe dysplasia or beginning invasion]. *Gynäkologe* 14 (4), 229-232.
166. Kataja V., Syrjanen S., Mantyarvi R., Yliskoski M., Saarikoski S., Syrjanen K. (1992). Prognostic factors in cervical human papillomavirus infections. *Sex TransmDis* 19 (3), 154-160.
167. Kavungu P., Ebeling K., Ebert S. (1983). [The effect of various factors on participation in cytologic screening in the prevention and early detection of cervical cancer]. *Z Ärztl Fortbild Jena* 77 (21), 883-885.

168. Kesic V. I., Soutter W. P., Sulovic V., Juznic N., Aleksic M., Ljubic A. (1993). A comparison of cytology and cervicography in cervical screening. *International Journal of Gynecological Cancer* 3 (6), 395-398.
169. Kestler H. A., Schule M., Schwenker F., Neumann H., Mattfeldt T. (1999). Neural classification of cytological smears from the cervix. *BIOMED TECH Bio-medizinische Technik* 44, 1-2.
170. Keyhani-Rofagha S., Palma T., O'Toole R. V. (1996). Automated screening for quality control using PAPNET: a study of 638 negative Pap smears. *Diagn Cytopathol* 14 (4), 316-320.
171. Khalbuss W. E., Rudomina D., Kauff N. D., Chuang L., Melamed M. R. (2000). SpinThin, a simple, inexpensive technique for preparation of thin-layer cervical cytology from liquid-based specimens: data on 791 cases. *Cancer* 90 (3), 135-142.
172. Kharazi M., Keyhani-Rofagha S., O'Toole R. (1993). Automated screening of vaginal cytology using PAPNET: comparative cytologic findings with biopsy. *Am J Clin Pathol* 99, 337.
173. Kim J. J., Wright T. C., Goldie S. J. (2002). Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined significance. *JAMA* 287 (18), 2382-2390.
174. Kisner H. J. (1994). Cytoc Corporation ThinPrep processor. *Clin Lab Manage Rev* 8, 154-160.
175. Kiviat N. (1996). Natural history of cervical neoplasia: overview and update. *Am J Obstet Gynecol* 175 (4 Pt 2), 1099-1104.
176. Knowles K., Bur M., Otis C. (1992). Comparison between conventional Papanicolaou smears and ThinPrep preparations for evaluation of cervicovaginal cytology. *Acta Cytol* 36, 582.
177. Kok M. R., Boon M. E. (1996). Consequences of neural network technology for cervical screening: increase in diagnostic consistency and positive scores. *Cancer* 78 (1), 112-117.
178. Kok M. R., Habers M. A., Schreiner-Kok P. G., Boon M. E. (1998). New paradigm for ASCUS diagnosis using neural networks. *Diagn Cytopathol* 19 (5), 361-366.
179. Kolles H., Stegmaier C., von Seebach H. B., Ziegler H. (1989). [Descriptive epidemiology and prognosis of malignant gynecologic tumors]. *Geburtshilfe und Frauenheilkunde* 49 (6), 573-578.
180. Koopmanschap (1990). Cervical-cancer screening: Attendance and cost-effectiveness. *Int J Cancer* 45, 410-5.

181. Koroschetz F. (1991). [Suspicious findings in gynecologic cytodiagnosis]. *Wien Klin Wochenschr* 103 (12), 357-61.
182. Koss (2000). Utility of liquid-based cytology for cervical carcinoma screening. *Cancer Cytopathology* 90, 67-8.
183. Koss L. G., Lin E., Schreiber K., Elgert P., Mango L. (1994). Evaluation of the PAPNET cytologic screening system for quality control of cervical smears. *Am J Clin Pathol* 101 (2), 220-229.
184. Koss L. G., Sherman M. E., Cohen M. B., Anes A. R., Darragh T. M., Lemos L. B., McClellan B. J., Rosenthal D. L., Keyhani-Rofagha S., Schreiber K., Valente P. T. (1997). Significant reduction in the rate of false-negative cervical smears with neural network-based technology (PAPNET Testing System). *Hum Pathol* 28 (10), 1196-1203.
185. Koutsky L. (1997). Epidemiology of genital human papillomavirus infection. *Am J Med* 102 (5a), 3-8.
186. Kraus H., Schneider A. (1984). [Frequency, causes and results of dubious smear findings (Papanicolaou III group)]. *Geburtshilfe-Frauenheilkd* 44 (2), 87-90.
187. Kreienberg R., Mobus V. (1995). Screening in cervical and endometrial cancer. *Onkologie* 3, 41-45.
188. Kreuger F. A., van Ballegooijen M., Doornewaard H. (1999). Is PAPNET suitable for primary screening? *Lancet* 353 (9162), 1374-1375.
189. Kunz J., Rondez R., Yoshizaki C., Fivian M., Held G., Lind B. (1998). [Comparison of conventional PAP smears with thin layer specimen (liquid-based PAP test) and correlation with cytopathological findings with HPV status using the hybrid capture system]. *Schweiz-Rundsch-Med-Prax* 87 (43), 1434-40.
190. Küppers V., Schnurch H. G. (1997). Current state of early diagnosis of female genital cancer. *Gynäkologie* 30 (8), 624-630.
191. Kurtycz D. F. I., Hoerl H. D. (2000). Thin-Layer Technology: Tempered Enthusiasm. *Diagnostic Cytopathology* 23 (1), 1-5.
192. Lasnik E., Breitenacker G. (1987). [Cytologic anamnesis of patients with invasive cervix cancer]. *Geburtshilfe-Frauenheilkd* 47 (8), 555-7.
193. Lavery C. R., Farnsworth A., Thurloe J. K., Grieves A., Bowditch R. (1997). Evaluation of the CytoRich slide preparation process. *Anal Quant Cytol Histol* 19 (3), 239-245.
194. Lavery C. R., Thurloe J. K., Redman N. L., Farnsworth A. (1995). An Australian trial of ThinPrep: a new cytopreparatory technique. *Cytopathology* 6 (3), 140-148.

195. Lee J. S., Kuan L., Oh S., Patten F. W., Wilbur D. C. (1998). A feasibility study of the AutoPap system location-guided screening. *Acta Cytol* 42 (1), 221-226.
196. Lee K. R., Ashfaq R., Birdsong G. G., Corkill M. E., McIntosh K. M., Inhorn S. L. (1997). Comparison of conventional Papanicolaou smears and a fluid-based, thin-layer system for cervical cancer screening. *ObstetGynecol* 90 (2), 278-284.
197. Leidl R. (1998). Der Effizienz auf der Spur: Eine Einführung in die ökonomische Evaluation. In *Das Public Health Buch. Gesundheit und Gesundheitswesen*, Schwarz FW, Badura B, Leidl R, Raspe H, Siegrist J, eds. Urban & Schwarzenberg, München.
198. Leidl R., von der Schulenburg J.-M., Wasem J. (1999). Ansätze und Methoden der ökonomischen Evaluation - eine internationale Perspektive, Vol 9. Nomos, Baden-Baden.
199. Lerma E., Colomo L., Carreras A., Esteva E., Quilez M., Prat J. (1998). Re-screening of atypical cervicovaginal smears using PAPNET. *Cancer* 84 (6), 361-365.
200. Lieu D. (1996). The Papanicolaou smear: Its value and limitations. *Journal of Family Practice* 42 (4), 391-399.
201. Linder J. (1997). Automation of the Papanicolaou smear: a technology assessment perspective. *Arch Pathol Lab Med* 121 (3), 282-6.
202. Linder J. (1998). Recent advances in thin-layer cytology. *Diagn Cytopathol* 18 (1), 24-32.
203. Linder J., Hutchinson M., Zahniser D. (1997). Improved performance of the Thinprep 2000 System in the detection of cervical disease: Comparison with the Thinprep Beta. *Laboratory Investigation* 76 (1), 194.
204. Linder J., Zahniser D. (1997). The ThinPrep Pap test. A review of clinical studies. *Acta Cytol* 41 (1), 30-38.
205. Linder J., Zahniser D. (1998). ThinPrep Papanicolaou testing to reduce false-negative cervical cytology. *Arch Pathol Lab Med* 122 (2), 139-144.
206. Lonky N. M., Sadeghi M., Tsadik G. W., Petitti D. (1999). The clinical significance of the poor correlation of cervical dysplasia and cervical malignancy with referral cytologic results. *American Journal of Obstetrics & Gynecology* 181 (3), 560-566.
207. Ludwig H. (1989). [Cervix cancer: incidence, diagnosis, therapeutic principle]. *Ther-Umsch* 46 (12), 855-63.

208. Lühmann D., Siebert U., Droste S., Von der Leyen U., Perleth M. (1999). Konzeption zur Anpassung und Erstellung von Technologiebewertungen in der deutschen Arbeitsgruppe Verfahrensbewertung in der Medizin. In Toolkit - Informationsmaterial für Verfasser von HTA-Berichten, Care GSWGTAfH, ed. German Scientific Working Group Technology Assessment for Health Care.
209. Mandelblatt J., Freeman H., Winczewski D., Cagney K., Williams S., Trowers R., Tang J., Kerner J. (1996). Implementation of a breast and cervical cancer screening program in a public hospital emergency department. *Annals of Emergency Medicine* 28 (5), 493-498.
210. Mandelblatt J. S., Fahs M. C. (1988). The cost-effectiveness of cervical cancer screening for low-income elderly women. *JAMA* 259 (16), 2409-2413.
211. Mandelblatt J. S., Lawrence W. F., Womack S. M., Jacobson D., Yi B., Hwang Y., Gold K., Barter J., Shah K. (2002). Benefits and Costs of Using HPV Testing to Screen for Cervical Cancer. *JAMA* 287 (18), 2372-2381.
212. Mango L. J. (1994). Computer-assisted cervical cancer screening using neural networks. *Cancer Letters* 77 (2-3), 155-162.
213. Mango L. J. (1996a). The FDA review process: obtaining premarket approval for the PAPNET Testing System. *Acta Cytol* 40 (1), 138-140.
214. Mango L. J. (1996b). Neuromedical Systems, Inc. *Acta Cytol* 40 (1), 53-9.
215. Mango L. J. (1997). Clinical validation of interactive cytologic screening. Automating the search, not the interpretation. *Acta Cytol* 41 (1), 93-97.
216. Mango L. J., Radensky P. W. (1998). Interactive neural network-assisted screening: A clinical assessment. *Acta Cytol* 42 (1), 233-245.
217. Mango L. J., Valente P. T. (1998). Neural-network-assisted analysis and microscopic rescreening in presumed negative cervical cytologic smears. A comparison. *Acta Cytol* 42 (1), 227-232.
218. Marshall C. J., Rowe L., Bentz J. S. (1999). Improved quality-control detection of false-negative Pap smears using the Autopap 300 QC system. *DiagnCytopathol* 20 (3), 170-174.
219. Matsunaga (1997). Cost-effectiveness analysis of mass screening for cervical cancer in Japan. *J Epidemiol* 7, 135-41.
220. Mayeaux Jr E. J., Harper M. B., Abreo F., Pope J. B., Phillips G. S. (1995). A comparison of the reliability of repeat cervical smears and colposcopy in patients with abnormal cervical cytology. *Journal of Family Practice* 40 (1), 57-62.
221. McCrory D. C., Matchar D. B. (1999). Evaluation of cervical cytology - systematic review. In <http://www.ahrq.gov>: Agency for Health Care Research and Quality (AHRQ).

222. McGoogan E., Colgan T. J., Ramzy I., Cochand-Priollet B., Davey D. D., Grohs H. K., Gurley A. M., Husain O. A., Hutchinson M. L., Knesel E. A., et al. (1998). Cell preparation methods and criteria for sample adequacy. International Academy of Cytology Task Force summary. *Diagnostic Cytology Towards the 21st Century: An International Expert Conference and Tutorial*. *Acta Cytol* 42 (1), 25-32.
223. McGoogan E., Reith A. (1996). Would monolayers provide more representative samples and improved preparations for cervical screening? Overview and evaluation of systems available. *Acta Cytol* 40 (1), 107-119.
224. McQuarrie H. G., Ogden J., Costa M. (2000). Understanding the financial impact of covering new screening technologies. The case of automated Pap smears. *J Reprod Med* 45 (11), 898-906.
225. Meerding W. J., Doornewaard H., van Ballegooijen M., Bos A., van der Graaf Y., van den Tweel J. G., van der Schouw Y. T., Habbema J. D. F. (2001). Cost analysis of PAPNET-assisted vs. conventional pap smear evaluation in primary screening of cervical smears. *Acta Cytol* 45 (1), 28-35.
226. Melamed M. R., Kamentsky L. A. (1969). An assessment of the potential role of automatic devices in cytology screening. *Obstet Gynecol Surv* 24 (7), 914-26.
227. Melnikow J., Nuovo J. (1996). Management choices for patients with squamous atypia on Papanicolaou smear. A toss up? *Med Care* 34, 336-47.
228. Melnikow J., Nuovo J. (1999). Reducing mortality due to cervical cancer. PAPNET fails the test. *Arch Fam Med* 8 (1), 56-57.
229. Melnikow J., Nuovo J., Paliescheskey M., Stewart G. K., Howell L., Green W. (1997). Detection of high-grade cervical dysplasia: impact of age and Bethesda system terminology. *Diagn Cytopathol* 17 (5), 321-325.
230. Meulman J. J., Zeppa P., Boon M. E., Rietveld W. J. (1992). Prediction of various grades of cervical neoplasia on plastic-embedded cytobrush samples. Discriminant analysis with qualitative and quantitative predictors. *Anal Quant Cytol Histol* 14 (1), 60-72.
231. Michelow P. M., Hlongwane N. F., Leiman G. (1997). Simulation of primary cervical cancer screening by the PAPNET system in an unscreened, high-risk community. *Acta Cytol* 41 (1), 88-92.
232. Minge L., Fleming M., VanGeem T., Bishop J. W. (2000). AutoCyte Prep system vs. conventional cervical cytology. Comparison based on 2,156 cases. *J Reprod Med* 45 (3), 179-184.
233. Mitchell H., Medley G. (1998). Detection of unsuspected abnormalities by PAPNET-assisted review. *Acta Cytol* 42 (1), 260-264.

234. Montz F. J., Farber F. L., Bristow R. E., Cornelison T. (2001). Impact of Increasing Papanicolaou Test Sensitivity and Compliance: A Modeled Cost and Outcomes Analysis. *Obstet Gynecol* 97, 781-788.
235. Montz F. J., Monk B. J., Fowler J. M., Nguyen L. (1992). Natural history of the minimally abnormal Papanicolaou smear. *Obstetrics & Gynecology* 80 (3 I), 385-388.
236. Muller C., Mandelblatt J. (1990). Costs and effectiveness of cervical cancer screening in elderly women. Washington, DC: US congress, Office of Technology Assessment.
237. Munck C. (1985). [Significance of intracervical cytology in the postmenopause]. *Zentralbl Gynakol* 107 (21), 1314-8.
238. Munoz N., Kato I., Bosch F. X. (1996). Risk factors for HPV DNA detection in middle-aged women. *Sex Transm Dis* 23, 504-510.
239. Myers E. R., McCrory D. C., Nanda K., Bastian L., Matchar D. B. (2000a). Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. *American Journal of Epidemiology* 151 (12), 1158-1171.
240. Myers E. R., McCrory D. C., Subramanian S., McCall N., Nanda K., Datta S., Matchar D. B. (2000b). Setting the target for a better cervical screening test: Characteristics of a cost-effective test for cervical neoplasia screening. *Obstetrics & Gynecology* 96 (5), 645-652.
241. Nagy P., Keller G. (1989). Correlation between cervicovaginal cytology and histology in endometrial carcinoma. *Zentralblatt für Gynäkologie* 111 (18), 1241-1245.
242. Nanda K., McCrory D. C., Myers E. R., Bastian L. A., Hasselblad V., Hickey J. D. (2000). Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med* 6 (132 (10)), 810-819.
243. Neumann G. (1988). [Mortality of malignant neoplasms of the uterus. An epidemiologic study]. *Geburtshilfe-Frauenheilkd* 48 (2), 65-71.
244. Neumann H. G., Barten G., Barten M., Schroder M. (1989). The frequency of suspicious and positive results of gynecocytological examinations in young women. *Zentralblatt für Gynäkologie* 111 (14), 947-951.
245. Neumann H. G., Seidenschnur G. (1987). [Computer-assisted screening for cervical carcinoma--the results for a 10-year period]. *Arch Geschwulstforsch* 57 (2), 129-39.
246. NHS (1999). Management of gynaecological cancers .Effective Health Care. (<http://www.york.ac.uk/inst/crd/ehc53htm>) 5 (3), 12.

247. Noorani H. Z., Arratoon C., Hall A. (1997). Assessment of techniques for cervical cancer screening. - nonsystematic review.: Canadian Coordinating Office for Health Technology Assessment (CCOHTA), pp. 1-33.
248. Nyirjesy I., Billingsley F. S., Forman M. R. (1998). Evaluation of atypical and low-grade cervical cytology in private practice. *Obstetrics & Gynecology* 92 (4 I), 601-607.
249. O'Brien B. (1997). A Tale of Two (or More) Cities: Geographic Transferability of Pharmacoeconomic Data. *Am Journal Managed Care* 3 (Suppl.), 33-39.
250. O'Leary T. J., Tellado M., Buckner S. B., Ali I. S., Stevens A., Ollayos C. W. (1998). PAPNET-assisted rescreening of cervical smears: cost and accuracy compared with a 100% manual rescreening strategy. *JAMA* 279 (3), 235-237.
251. Oortmarssen G., Habbema J. (1991). Epidemiological evidence for age-dependant regression of pre-invasive cervical cancer. *British Journal of Cancer* 64, 559-565.
252. Östör A. G. (1993). Natural history of cervical intraepithelial neoplasia: a critical review. *Int J Gynecol Pathol* 12 (2), 186-192.
253. Ouwerkerk-Noordam E., Boon M. E., Beck S. (1994). Computer-assisted primary screening of cervical smears using the PAPNET method: comparison with conventional screening and evaluation of the role of the cytologist. *Cytopathology* 5 (4), 211-218.
254. Pallatroni L. (1999). The new technology for cervical cancer screening: costs versus reimbursement. *MLO Med Lab Obs* 31 (3), 44-6, 48, 50.
255. Papillo J. L., Zarka M. A., St John T. L. (1998). Evaluation of the ThinPrep Pap test in clinical practice. A seven-month, 16,314-case experience in northern Vermont. *Acta Cytol* 42 (1), 203-208.
256. Park I. A., Lee S. N., Chae S. W., Park K. H., Kim J. W., Lee H. P. (2001). Comparing the accuracy of ThinPrep Pap tests and conventional Papanicolaou smears on the basis of the histologic diagnosis: a clinical study of women with cervical abnormalities. *Acta Cytol* 45 (4), 525-531.
257. Paterok E. M., Siebzehrubl E., Meyer M. (1995). [Age specific and age corrected mortality in gynecological carcinoma--including breast carcinoma]. *Geburts-hilfe und Frauenheilkunde* 55 (4), 177-81.
258. Patten S., Lee J., Wilbur D., Bonfiglio T., Colgan T., Richart R., Taylor D., Cramer H., Moinuddin S. (1996a). Multicenter sensitivity studies of the Auto-Pap(R) 300QC for detection of abnormalities in cervical cytology specimens. *Laboratory Investigation* 74 (1), 194.
259. Patten S. F., Lee J. S., Nelson A. C. (1996b). NeoPath, Inc. NeoPath AutoPap 300 Automatic Pap Screener System. *Acta Cytol* 40 (1), 45-52.

260. Patten S. F., Lee J. S. J., Wilbur D. C., Bonfiglio T. A., Colgan T. J., Richart R. M., Cramer H., Moinuddin S. (1997). The AutoPap 300 QC system multicenter clinical trials for use in quality control rescreening of cervical smears - I. A prospective intended use study. *Cancer Cytopathology* 81 (6), 337-342.
261. Payne N., Chilcott J., McGoogan E. (2000). Liquid-based cytology in cervical screening: a rapid and systematic review. In <http://www.ncchta.org/execsumm/summ418.htm>: The National Coordinating Centre for Health Technology Assessment (NCCHTA), pp. 73.
262. Perlman S. E. (2000). Pap smears: screening, interpretation, treatment. [Review]. *Adolescent Medicine* 10 (2), 243-254.
263. Pete I., Toth V., Bosze P. (1998). The value of colposcopy in screening cervical carcinoma. *European Journal of Gynaecological Oncology* 19 (2), 120-122.
264. Petry K. U., Glaubitz M., Maschek H., Bohmer G., Linge G., Kuhnle H. (1996). Loop electrosurgical excisional procedure (LEEP) for the treatment of cervical neoplasia. *Geburtshilfe und Frauenheilkunde* 56 (10), 513-516.
265. Promer H., Heiss O., Staudach A., Thurner J. (1993). [Cervix cancer despite cytology. A new critical look at cervix cancer prevention]. *Gynäkol Geburtshilfliche Rundsch* 1, 269-270.
266. Quaas J., Heinrich J. (1998). [Screening for cervix carcinoma--a retrospective comparison between new and established federal German states]. *Zentralblatt für Gynäkologie* 120 (9), 437-443.
267. Quantin C., Dusserre L., Montaud A. M., Mottot C., Feldman J. P. (1992). A model for quality assessment in cervical cytology used as a screening test. *Qual Assur Health Care* 4 (2), 105-113.
268. Quitllet F. A., Morta M. C., Canas A., Castellanos E., Carreras R. (1997). Cytologic atypia: Clinical significance and follow-up recommendations. *Acta Cytol* 41 (2), 504-506.
269. Raab S. S. (1997). The cost-effectiveness of cervical-vaginal rescreening. *Am J Clin Pathol* 108 (5), 525-536.
270. Raab S. S., Bishop N. S., Zaleski M. S. (1999a). Cost effectiveness of rescreening cervicovaginal smears. *Am J Clin Pathol* 111 (5), 601-609.
271. Raab S. S., Zaleski M. S., Silverman J. F. (1999b). The cost-effectiveness of the cytology laboratory and new cytology technologies in cervical cancer prevention. *Am J Clin Pathol* 111 (2), 259-266.
272. Radensky P. W., Mango L. J. (1998). Interactive neural-network-assisted screening. An economic assessment. *Acta Cytol* 42 (1), 246-252.

273. Richart R. M., Patten S. F., Jr., Lee J. S. (1996). Prospects for automated cytology. *Obstetrics & Gynecology Clinics of North America* 23 (4), 853-859.
274. Roberts J. M., Gurley A. M., Thurloe J. K., Bowditch R., Lavery C. R. (1997). Evaluation of the ThinPrep Pap test as an adjunct to the conventional Pap smear. *Med J Aust* 167 (9), 466-469.
275. Roberts J. M., Thurloe J. K., Bowditch R. C., Humcevic J., Lavery C. R. (1999). Comparison of ThinPrep and Pap smear in relation to prediction of adenocarcinoma in situ. *Acta Cytol* 43 (1), 74-80.
276. Robra B. P., Dierks M. L. (1990). [Development of the participation in early cancer screening of women]. *Gynäkologe* 23 (6), 308-11.
277. Rodriguez Castro R., Garcia Gonzales R. (1978). [Cytohistologic correlation of stage III using Papanicolaou's method]. *Zentralbl Gynäkol* 100 (22), 1499-1500.
278. Ronco G., Montanari G., Aimone V., Parisio F., Segnan N., Valle A., Volante R. (1996). Estimating the sensitivity of cervical cytology: errors of interpretation and test limitations. *Cytopathology* 7 (3), 151-158.
279. Rosenthal D. L. (1998). Automation and the endangered future of the Pap test. *Journal of the National Cancer Institute* 90, 738-49.
280. Rosenthal D. L., Acosta D., Peters R. K. (1996). Computer-assisted rescreening of clinically important false negative cervical smears using the PAPNET Testing System. *Acta Cytol* 40 (1), 120-126.
281. Rummel H. H., Fick R., Heberling D., Schubert D. (1977). [Cytologic follow-up examination of patients with a suspicious Papanicolaou type 3D smear]. *Geburtshilfe-Frauenheilkd* 37 (6), 521-526.
282. Sasieni P. D. (1999). Statistical analysis of the performance of diagnostic tests. *Cytopathology* 10 (2), 73-78.
283. Sato S., Matunaga G., Tsuji I., Yajima A., Sasaki H. (1999). Determining the cost-effectiveness of mass screening for cervical cancer using common analytic models. *Acta Cytol* 43 (6), 1006-1014.
284. Sawaya G. F., Brown A. D., Washington A. E., Garber A. M. (2001). Current Approaches to Cervical-Cancer Screening. *NEJM* 344 (21), 1603-1607.
285. Sawaya G. F., Grimes D. A. (1999). New technologies in cervical cytology screening: a word of caution. *Obstet Gynecol* 94 (2), 307-10.
286. Sawaya G. F., Washington A. E. (1999). Cervical cancer screening: Which techniques should be used and why? *Clinical Obstetrics and Gynecology* 42 (4), 922-938.
287. Schechter C. B. (1996). Cost-effectiveness of rescreening conventionally prepared cervical smears by PAPNET testing. *Acta Cytol* 40 (6), 1272-1282.

288. Schenck U. (2000). Cytological screening program - Seriously challenged by new methods? *Geburtshilfe und Frauenheilkunde* 60 (9), M125-m129.
289. Schenck U., von Karsa L. (2000). Cervical cancer screening in Germany. *European Journal of Cancer* 36 (17), 2221-6.
290. Schiffman M. H., Bauer H. M., Hoover R. N., Glass A. G., Cadell D. M., Rush B. B., Scott D. R., Sherman M. E., Kurman R. J., Wacholder S., et al. (1993). Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *Journal of the National Cancer Institute* 85 (12), 958-964.
291. Schlegel H., Kuhndel K. (1983). [Diagnostic verification of Group III and III D cytological tests with special reference to follow-up]. *Zentralbl Gynäkol* 105 (24), 1585-1596.
292. Schneider, al. e. (2001). *Scriptum für die Hauptvorlesung "Frauenheilkunde"*. (<http://www.uni-jenade/ufk/cd/folder/starthtm>), 24.
293. Schneider A., Durst M., Jochmus I., Gissmann L. (1999). Epidemiology, aetiology and prevention of carcinoma of the cervix. *Gynäkologe* 32 (4), 247-260.
294. Schneider A., Hoyer H., Lotz B., Leistritz S., Kuhne-Heid R., Nindl I., Müller B., Haerting J., Durst M. (2000). Screening for high-grade cervical intra-epithelial neoplasia and cancer by testing for high-risk HPV, routine cytology or colposcopy. *International Journal of Cancer* 89 (6), 529-534.
295. Schneider A., Shah K. (1989). The role of vitamins in the etiology of cervical neoplasia: an epidemiological review. *Arch Gynecol Obstet* 246 (1), 1-13.
296. Schneider A., Zahm D. M. (1996). New adjunctive methods for cervical cancer screening. *Obstetrics & Gynecology Clinics of North America* 23 (3), 657-673.
297. Schneider A., Zahm D. M., Kirchmayr R., Schneider V. L. (1996). Screening for cervical intraepithelial neoplasia grade 2/3: Validity of cytologic study, cervicography, and human papillomavirus detection. *Am J Obstet Gynecol* 174 (5), 1534-1541.
298. Sheets E. E., Constantineee N. M., Dinisco S., Dean B., Cibas E. S. (1995). Colposcopically Directed Biopsies Provide a Basis for Comparing the Accuracy of ThinPrep and Papanicolaou smears. *Journal of Gynecologic Techniques* 1, 27-33.
299. Sherlaw-Johnson C., Gallivan S., Jenkins D., Jones M. H. (1994). Cytological screening and management of abnormalities in prevention of cervical cancer: an overview with stochastic modelling. *J Clin Pathol* 47 (5), 430-435.

300. Sherman M. E., Mango L. J., Kelly D., Paull G., Ludin V., Copeland C., Solomon D., Schiffman M. H. (1994). PAPNET analysis of reportedly negative smears preceding the diagnosis of a high-grade squamous intraepithelial lesion or carcinoma. *Mod Pathol* 7 (5), 578-581.
301. Sherman M. E., Schiffman M., Herrero R., Kelly D., Bratti C., Mango L. J., Alfaro M., Hutchinson M. L., Mena F., Hildesheim A., et al. (1998). Performance of a semiautomated Papanicolaou smear screening system: results of a population-based study conducted in Guanacaste, Costa Rica. *Cancer* 84 (5), 273-280.
302. Sherman M. E., Schiffman M. H., Mango L. J., Kelly D., Acosta D., Cason Z., Elgert P., Zaleski S., Scott D. R., Kurman R. J., et al. (1997). Evaluation of PAPNET testing as an ancillary tool to clarify the status of the "atypical" cervical smear. *Mod Pathol* 10 (6), 564-571.
303. Shield P. W., Cox N. C. (1998). The sensitivity of rapid (partial) review of cervical smears. *Cytopathology* 9 (2), 84-92.
304. Siebert U. (2002a). Decision-Analytic Modelling in Combination with Competence Network Collaboration Provides the Opportunity to Transfer Evidence from Other Countries to the National Context. Paper presented at: Proceedings of the 18th Annual Meeting of the International Society of Technology Assessment in Health Care (ISTAHC). Berlin, Urban & Fischer.
305. Siebert U. (2002b). Entscheidungsanalytische Modelle zur Sicherung der Übertragbarkeit internationaler Evidenz aus HTA auf den Kontext des deutschen Gesundheitssystems.: HTA-Methodenreport im Auftrag der Deutschen Agentur für Health Technology Assessment (DIMDI / Bundesministerium für Gesundheit). Zur Veröffentlichung eingereicht.
306. Siebert U., Behrend C., Mühlberger N., Wasem J., Greiner W., von der Schulenburg J. M., Welte R., Leidl R. (1999). Entwicklung eines Kriterienkataloges zur Beschreibung und Bewertung ökonomischer Evaluationsstudien in Deutschland. In Ansätze und Methoden der ökonomischen Evaluation - eine internationale Perspektive. Health Technology Assessment, Leidl R, von der Schulenburg, J.M., Wasem, J., ed. Nomos, Baden-Baden, pp. 156-170.
307. Siebert U., Kurth T. (2000). Lebensqualität als Parameter von medizinischen Entscheidungsanalysen. In Lebensqualitätsforschung und Gesundheitsökonomie in der Medizin, Ravens-Sieberer U, Cieza A, von Steinbüchel N, Bullinger M, eds. Ecomed, Landsberg.
308. Sigurdsson K. (1995). Quality assurance in cervical cancer screening: The Icelandic experience 1964-1993. *European Journal of Cancer* 31 (5), 728-734.
309. Slawson D. C., Bennett J. H., Simon L. J., Herman J. M. (1994). Should all women with cervical atypia be referred for colposcopy: a HARNET study. Harrisburgh Area Research Network. *Journal of Family Practice* 38 (4), 387-392.

310. Sloan F. A. (1995). *Valuing Health Care, Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies*. Cambridge University Press, Cambridge.
311. Smith B. L., Lee M., Leader S., Wertlake P. (1999). Economic impact of automated primary screening for cervical cancer. *J Reprod Med* 44 (6), 518-528.
312. Smith B. L., Lee M., Leader S., Wertlake P. (2000). Economic impact of automated primary screening for cervical cancer. *J Reprod Med* 45, 83-84.
313. Solomon D., Schiffman M., Tarone R. (2001). Comparison of Three Management Strategies for Patients With Atypical Squamous Cells of Undetermined Significance: Baseline Results >From a Randomized Trial. *J Natl Cancer Inst* 93 (4), 293-299.
314. Solomon H. M., Frist S. (1998). PAPNET testing for HSILs. The few cell/small cell challenge. *Acta Cytol* 42 (1), 253-259.
315. Sommer H., Noschel H., Stech D. (1992). [Compliance in gynecologic screening and causes for delayed tumor diagnosis in East Thuringen]. *Zentralbl Gynäkol* 114 (6), 312-315.
316. Sonnenberg F. A., Beck J. R. (1993). Markov models in medical decision making: a practical guide. *Med Decis Making* 13 (4), 322-38.
317. Soost H. J., Lange H. J., Lehmacher W., Ruffing-Kullmann B. (1991). The validation of cervical cytology. Sensitivity, specificity and predictive values. *Acta Cytol* 35 (1), 8-14.
318. Soost H. J., Lehmacher W., Ruffing Kullmann B. (1989). How far have preliminary-care examinations helped to recognise carcinoma of the cervix uteri well in time? *Geburtshilfe und Frauenheilkunde* 49 (6), 579-586.
319. Spitzer M. (1998). Cervical screening adjuncts: recent advances. *Am J Obstet Gynecol* 179 (2), 544-556.
320. Sprenger (1996). The false negative rate in cervical cytology. Comparison of monolayers to conventional smears. *Acta Cytol* 40, 81-9.
321. Steps H. (1982). [Positive cytological smears in adolescents and young women]. *Geburtshilfe-Frauenheilkd* 42 (8), 613-615.
322. Stevens M. W., Milne A. J., James K. A., Brancheau D., Ellison D., Kuan L. (1997a). Effectiveness of automated cervical cytology rescreening using the AutoPap 300 QC System. *Diagn Cytopathol* 16 (6), 505-512.
323. Stevens S. A., Cockburn J., Hirst S., Jolley D. (1997b). An evaluation of educational outreach to general practitioners as part of a statewide cervical screening program. *Am J Public Health* 87 (7), 1177-1181.

324. Stoler M. H. (2000). Advances in cervical screening technology. *Modern Pathology* 13 (3), 275-284.
325. Stoler M. H., Schiffman M. (2001). Interobserver Reproducibility of Cervical Cytologic and Histologic Interpretations Realistic Estimates From the ASCUS-LSIL Triage Study. *JAMA* 285 (11), 1500-1505.
326. Stoll W. (1989). [The gynecologic check-up]. *Ther-Umsch* 46 (6), 361-7.
327. Syrjanen K., Kataja V., Yliskoski M. (1992). Natural history of cervical humanpapillomavirus lesions does not substantiate the biologic relevance of the Bethesda System. *Obstet Gynecol* 79, 675-682.
328. Tawa K., Forsythe A., Cove J. K., Saltz A., Peters H. W., Watring W. G. (1988). A comparison of the Papanicolaou smear and the cervigram: Sensitivity, specificity, and cost analysis. *Obstetrics & Gynecology* 71 (2), 229-235.
329. Taylor L. A., Sorensen S. V., Ray N. F., Halpern M. T., Harper D. M. (2000). Cost-effectiveness of the conventional papanicolaou test with a new adjunct to cytological screening for squamous cell carcinoma of the uterine cervix and its precursors. *Arch Fam Med* 9 (8), 713-721.
330. Troni G. M., Cipparrone I., Cariaggi M. P., Ciatto S., Miccinesi G., Zappa M., Confortini M. (2000). Detection of false-negative pap smears using the PAPNET system. *Tumori* 86 (6), 455-457.
331. Tumorzentrum München (1999). Tumorzentrum München: Empfehlungen zur Diagnostik, Therapie und Nachsorge. Zervixkarzinom. Tumorzentrum München.
332. van Ballegooijen M., van den Akker-van Marle M. E., Patnick J., Lynge E., Arbyn M., Anttila A., Ronco G., Dik J., Habbema F., Bos A. B., et al. (2000). Overview of important cervical cancer screening process values in european union (EU) countries, and tentative predictions of the corresponding effectiveness and cost-effectiveness. *Europ J Cancer* 36 (17), 2177-2188.
333. Van den Akker-van Marle M. E., van Ballegooijen M., Boer R., van Oortmarssen G. J., Habbema J. D. F., Sherlaw-Johnson C., Gallivan S., Jenkins D. (1999). Withdrawing low risk women from cervical screening programmes (multiple letters) [15]. *BMJ* 318 (7201), 58.
334. Van Duyne W. M. J., Appelman C., Collette H. J. A., Drogendijk A. C., van der G. Y., Habbema J. D. F., Lee-Bylisma A., Pal R. S., Van Veen W. A., Verbrugge H. P., et al. (1989). Population screening for cervical cancer in the Netherlands. *International Journal of Epidemiology* 18 (4), 775-781.
335. Vassilakos P., Saurel J., Rondez R. (1999). Direct-to-vial use of the AutoCyte PREP liquid-based preparation for cervical-vaginal specimens in three European laboratories. *Acta Cytol* 43 (1), 65-68.

336. Vassilakos P., Schwartz D., de Marval F., Yousfi L., Broquet G., Mathez-Loic F., Campana A., Major A. (2000). Biopsy-based comparison of liquid-based, thin-layer preparations to conventional Pap smears. *J ReprodMed* 45 (1), 11-16.
337. Vessey M. P., Villard-Mackintosh L. (1992). The epidemiology of hysterectomy: findings in a large cohort study. *British Journal of Obstetrics and Gynaecology* 99, 402-407.
338. Wagner J. L. (1997). Cost-effectiveness of screening for common cancers. *Cancer and Metastasis Reviews* 16 (3-4), 281-294.
339. Wasem J. (1999). QALY-League Tables - ein Weg zu mehr Ergebnisorientierung im Gesundheitswesen? In *Das Gesundheitswesen in Deutschland: Von der Kosten- zur Nutzenorientierung. Gesundheitswissenschaftliche Analysen*, Laaser U, Schwalbe A, eds. Verlag Hans Jacobs, Lage, pp. 105-115.
340. Wasem J., Siebert U. (1999). Gesundheitsökonomische Parameter einer Evidence-based medicine. *Z Ärztl Fortbild Qualitätssich* 93 (6), 427-36.
341. Weinstein M. C. (1999). High-priced technology can be good value for money. *Annals of Internal Medicine* 130 (10), 857-8.
342. Weintraub J., Morabia A. (2000). Efficacy of a liquid-based thin layer method for cervical cancer screening in a population with a low incidence of cervical cancer. *DiagnCytopathol* 22 (1), 52-59.
343. Welte R., Leidl R. (1998). Übertragung ausländischer Studienergebnisse aus dem Bereich der ökonomischen Evaluation auf Deutschland: Probleme und Lösungsvorschläge. Projekt "Aufbau einer Datenbasis 'Evaluation medizinischer Verfahren und Technologien' in der Bundesrepublik". ([http://wwwdimdide](http://www.dimdide)).
344. Wertlake P. (1999). Results of AutoPap system-assisted and manual cytologic screening. A comparison. *J ReprodMed* 44 (1), 11-17.
345. Widra E. A., Dookhan D., Jordan A., McCue P., Bibbo M., Dunton C. J. (1994). Evaluation of the atypical cytologic smear. Validity of the 1991 Bethesda System. *Journal of Reproductive Medicine* 39 (9), 682-684.
346. Wilbur D., Dubeshter B., Angel C. (1996a). Use of thin-layer preparation for gynaecologic smears, with an emphasis on the cytomorphology of high grade intraepithelial lesions and carcinomas. *Diagn cytopathol* 14, 201-16.
347. Wilbur D. C., Bonfiglio T. A., Rutkowski M. A., Atkison K. M., Richart R. M., Lee J. S., Patten S. F. (1996b). Sensitivity of the AutoPap 300 QC System for cervical cytologic abnormalities. Biopsy data confirmation. *Acta Cytol* 40 (1), 127-132.
348. Wilbur D. C., Facik M. S., Rutkowski M. A., Mulford D. K., Atkison K. M. (1997). Clinical trials of the CytoRich specimen-preparation device for cervical cytology. Preliminary results. *Acta Cytol* 41 (1), 24-29.

349. Wilbur D. C., Prey M. U., Miller W. M., Pawlick G. F., Colgan T. J. (1998). The AutoPap system for primary screening in cervical cytology. Comparing the results of a prospective, intended-use study with routine manual practice. *Acta Cytol* 42 (1), 214-220.
350. Wilbur D. C., Prey M. U., Miller W. M., Pawlick G. F., Colgan T. J., Taylor D. D. (1999). Detection of high grade squamous intraepithelial lesions and tumors using the AutoPap System: results of a primary screening clinical trial. *Cancer* 87 (6), 354-358.
351. Wilson S., Woodman C. (1995). Assessing the effectiveness of cervical screening. *ClinObstetGynecol* 38 (3), 577-584.
352. Wong J. B., Sonnenberg F. A., Salem D. N., Pauker S. G. (1990). Myocardial revascularization for chronic stable angina. Analysis of the role of percutaneous transluminal coronary angioplasty based on data available in 1989. *Ann Intern Med* 113, 852-871.
353. Wunder G., Quaeitzsch U. (1973). [Significance and practical consequences of the suspicious colpocytologic smear (Papanicolaou group 3). Studied on about 140 000 remittances of the Cytologic Examination Center, Ludwigshafen on Rhine]. *Z-Allgemeinmed* 49 (30), 1472.
354. Zimmer S., Krahnert U., Munck C. (1987). [Significance of sources of error and quality control in cytologic mass screening]. *Zentralbl Gynäkol* 109 (4), 209-217.

III C 8. 2 Eingeschlossene Studien ökonomische Evaluation (Kapitel II)

1. AHTAC Review of automated and semi-automated cervical screening devices. 1998
2. Brown AD, Garber AM: Cost-effectiveness of 3 methods to enhance the sensitivity of Papanicolaou testing. In *Jama* 281 (1999) 4, 347-353.
3. Hutchinson ML, Farber FL: Cost-effectiveness of new cytology technologies. In *Am J Clin Pathol* 113 (2000) 1, 149-151.
4. McCrory DC, Matchar DB Evaluation of cervical cytology - systematic review. 1999
5. Montz FJ, Farber FL, Bristow RE, Cornelison T: Impact of Increasing Papanicolaou Test Sensitivity and Compliance: A Modeled Cost and Outcomes Analysis. In *Obstet Gynecol* 97 (2001) 781-788.

6. Noorani HZ, Menon D: Cost-effectiveness of rescreening strategies for cervical cancer [abstract]. In *AnnuMeetInt Soc Technol AssessHealth Care* 13 (1997) 135.
7. Payne N, Chilcott J, McGoogan E Liquid-based cytology in cervical screening: a rapid and systematic review. 2000
8. Raab SS, Zaleski MS, Silverman JF: The cost-effectiveness of the cytology laboratory and new cytology technologies in cervical cancer prevention. In *Am J Clin Pathol* 111 (1999) 2, 259-266.
9. Radensky PW, Mango LJ: Interactive neural-network-assisted screening. An economic assessment. In *Acta Cytol* 42 (1998) 1, 246-252.
10. Schechter CB: Cost-effectiveness of rescreening conventionally prepared cervical smears by PAPNET testing. In *Acta Cytol* 40 (1996) 6, 1272-1282.
11. Smith BL, Lee M, Leader S, Wertlake P: Economic impact of automated primary screening for cervical cancer. In *J Reprod Med* 44 (1999) 6, 518-528.

III C 8. 3 Ausgeschlossene Studien ökonomische Evaluation (Kapitel II)

1. Broadstock M Effectiveness and cost-effectiveness of automated and semi-automated cervical screening devices: a systematic review. 2000
2. Brotzman GL, Kretzchmar S, Ferguson D, Gottlieb M, Stowe C: Costs and outcomes of PAPNET secondary screening technology for cervical cytologic evaluation. A community hospital's experience. In *Arch Fam Med* 8 (1999) 1, 52-55.
3. Brown AD, Garber AM Cost-effectiveness of three new technologies to enhance Pap testing. 1998 Chicago.
4. Check W: Three approaches to improved screening. In *CAP Today* 11 (1997) 6, 44.
5. ECRI Automated monolayer slide preparation devices for Pap smear screening. 1998 Plymouth Meeting, PA.
6. ECRI Automated Pap smear rescreening technologies 1998 Plymouth Meeting, PA.
7. ECRI Automated monolayer slide preparation system for pap smear screening: ThinPrep 2000. 1999 Plymouth Meeting, PA.

8. ECRI Automated pap smear screening technologies: the autopap system. 2000 Plymouth Meeting, PA.
9. Hyde C, Cuzick J, Sasieni P Cervical screening programs: can automation help? Evidence from a systematic review and economic modelling, applied to the UK - secondary research (project). 1999
10. Kisner HJ: Cytoc Corporation ThinPrep processor. In Clin Lab Manage Rev 8 (1994) 154-160.
11. McQuarrie HG, Ogden J, Costa M: Understanding the financial impact of covering new screening technologies. The case of automated Pap smears. In J Reprod Med 45 (2000) 11, 898-906.
12. Meerding WJ, Doornewaard H, van Ballegooijen M, Bos A, van der Graaf Y, van den Tweel JG, van der Schouw YT, Habbema JDF: Cost analysis of PAPNET-assisted vs. conventional pap smear evaluation in primary screening of cervical smears. In Acta Cytol 45 (2001) 1, 28-35.
13. Myers ER, McCrory DC, Nanda K, Bastian L, Matchar DB: Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. In American Journal of Epidemiology 151 (2000) 12, 1158-1171.
14. O'Leary TJ, Tellado M, Buckner SB, Ali IS, Stevens A, Ollayos CW: PAPNET-assisted rescreening of cervical smears: cost and accuracy compared with a 100% manual rescreening strategy. In JAMA 279 (1998) 3, 235-237.
15. Pallatroni L: The new technology for cervical cancer screening: costs versus reimbursement. In MLO Med Lab Obs 31 (1999) 3, 44-46, 48, 50.
16. Troni GM, Cipparrone I, Cariaggi MP, Ciatto S, Miccinesi G, Zappa M, Confortini M: Detection of false-negative pap smears using the PAPNET system. In Tumori 86 (2000) 6, 455-457.

III C 8. 4 Im entscheidungsanalytischen Modell berücksichtigte Literatur (Kapitel III)

1. AHTAC Review of automated and semi-automated cervical screening devices. 1998
2. Anonymous: Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation - Revidierte Fassung des Hannoveraner Konsensus. Hannoveraner Konsensus Gruppe. In Gesundheitsökonomie und Qualitätsmanagement 4 (1999) A62-A65.

3. Anonymous Empfehlungen zur Diagnostik, Therapie und Nachsorge: Zervixkarzinom. 1999
4. Anonymous: Frauengesundheitsbericht Bremen 2001. In Der Senator für Arbeit, Frauen, Gesundheit, Jugend und Soziales. Abteilung Gesundheitswesen Referat Gesundheitsberichterstattung. <http://www.bremen.de/info/frauengesundheitsbericht/>, Freie Hansestadt Bremen, 2002, 57-65.
5. Arbeitsgemeinschaft bevölkerungsbezogener Krebsregister in Deutschland. (Verantwortliche Autoren: Batzler W.U. B-EC, Eisinger B., Lehnert M., Schön D., Schütz J., Stegmaier C.): Krebs in Deutschland - Häufigkeiten und Trends. In <http://www.rkide/KREBS> 3., aktualisierte Ausgabe. Saarbrücken. (2002)
6. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF): Kurzgefasste Interdisziplinäre Leitlinien 2000: Zervixkarzinom. In <http://www.uni-duesseldorf.de/WWW/AWMF/ziplinäre> (2000) 269ff.
7. Bishop JW, Bigner SH, Colgan TJ, Husain M, Howell LP, McIntosh KM, Taylor DA, Sadeghi MH: Multicenter masked evaluation of AutoCyte PREP thin layers with matched conventional smears. Including initial biopsy results. In *Acta Cytol* 42 (1998) 1, 189-197.
8. Bolick DR, Hellman DJ: Laboratory implementation and efficacy assessment of the ThinPrep cervical cancer screening system. In *Acta Cytol* 42 (1998) 1, 209-213.
9. Brenner H, Stegmaier C, Ziegler H: [Incidence of secondary tumors following cervical neoplasia in Saarland 1968-1987]. In *Geburtshilfe und Frauenheilkunde* 50 (1990) 8, 614-618.
10. Brenner H, Wiebelt H, Ziegler H: Progress in early detection of cervical carcinoma - Based on data of the cancer registry of Saarland. In *Geburtshilfe und Frauenheilkunde* 50 (1990) 4, 304-309.
11. Broadstock M Effectiveness and cost-effectiveness of automated and semi-automated cervical screening devices: a systematic review. 2000
12. Brown AD, Garber AM: Cost-effectiveness of 3 methods to enhance the sensitivity of Papanicolaou testing. In *Jama* 281 (1999) 4, 347-353.
13. Corkill M, Knapp D, Martin J, Hutchinson ML: Specimen adequacy of ThinPrep sample preparations in a direct-to-vial study. In *Acta Cytol* 41 (1997) 1, 39-44.
14. Eddy DM: Screening for cervical cancer. In *Ann Intern Med* 113 (1990) 3, 214-226.
15. Fahs MC, Mandelblatt J, Schechter C, Muller C: Cost effectiveness of cervical cancer screening for the elderly. In *Ann Intern Med* 117 (1992) 6, 520-527.

16. Gold MR, Siegel JE, Russell LB, Weinstein MC: Cost-Effectiveness in Health and Medicine. Oxford University Press Inc., New York, 1996,
17. Goldie S, Kuhn L, Denny L, Pollack A, Wright TC: Policy Analysis of Cervical Cancer Screening Strategies in Low-Resource Settings: Clinical Benefits and Cost-effectiveness. In *JAMA* 285 (2001) 24, 3107-3115.
18. Gustafsson L, Ponten J, Bergstrom R, Adami HO: International incidence rates of invasive cervical cancer before cytological screening. In *International Journal of Cancer* 71 (1997) 2, 159-165.
19. Hildesheim A, Schiffman M, Gravitt P: Persistence of type-specific human papillomavirus infection among cytologically normal women. In *J Infect Dis* 169 (1994) 235-240.
20. Ho GYF, Bierman R, Beardsley L: Natural history of cervicovaginal papillomavirus infection in young women. In *N Engl J Med* 338 (1998) 423-428.
21. Hutchinson ML, Farber FL: Cost-effectiveness of new cytology technologies. In *Am J Clin Pathol* 113 (2000) 1, 149-151.
22. Hutchinson ML, Zahniser DJ, Sherman ME, Herrero R, Alfaro M, Bratti MC, Hildesheim A, Lorincz AT, Greenberg MD, Morales J, Schiffman M: Utility of liquid-based cytology for cervical carcinoma screening: results of a population-based study conducted in a region of Costa Rica with a high incidence of cervical carcinoma. In *Cancer* 87 (1999) 2, 48-55.
23. Inhorn SL: Validation of the ThinPrep Papanicolaou test for cervical cancer diagnosis. In *J lower Genital Tract Disease* 2 (1998) 208-212.
24. Inhorn SL, Linder J, Wilbur DC, Zahniser D: Diagnosis of invasive cervical carcinoma with fluid-based, thin-layer slide preparation method. In 5 (1998) 4, 164.
25. Kataja V, Syrjanen S, Mantyjarvi R, Yliskoski M, Saarikoski S, Syrjanen K: Prognostic factors in cervical human papillomavirus infections. In *Sex TransmDis* 19 (1992) 3, 154-160.
26. Kim JJ, Wright TC, Goldie SJ: Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined significance. In *JAMA* 287 (2002) 18, 2382-2390.
27. Kiviat N: Natural history of cervical neoplasia: overview and update. In *Am J Obstet Gynecol* 175 (1996) 4 Pt 2, 1099-1104.
28. Koutsky L: Epidemiology of genital human papillomavirus infection. In *Am J Med* 102 (1997) 5a, 3-8.

29. Mandelblatt JS, Lawrence WF, Womack SM, Jacobson D, Yi B, Hwang Y, Gold K, Barter J, Shah K: Benefits and Costs of Using HPV Testing to Screen for Cervical Cancer. In *JAMA* 287 (2002) 18, 2372-2381.
30. McCrory DC, Matchar DB Evaluation of cervical cytology - systematic review. 1999
31. Minge L, Fleming M, VanGeem T, Bishop JW: AutoCyte Prep system vs. conventional cervical cytology. Comparison based on 2,156 cases. In *J Reprod Med* 45 (2000) 3, 179-184.
32. Munoz N, Kato I, Bosch FX: Risk factors for HPV DNA detection in middle-aged women. In *Sex Transm Dis* 23 (1996) 504-510.
33. Myers ER: Improving the sensitivity of cervical cytology: what are the issues? [editorial]. In *Am J ManagCare* 6 (2000) 7, 838-840.
34. Myers ER, McCrory DC, Nanda K, Bastian L, Matchar DB: Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. In *American Journal of Epidemiology* 151 (2000) 12, 1158-1171.
35. Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD: Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. In *Ann Intern Med* 6 (2000) 132 (10), 810-819.
36. Oortmarssen G, Habbema J: Epidemiological evidence for age-dependant regression of pre-invasive cervical cancer. In *British Journal of Cancer* 64 (1991) 559-565.
37. Östör AG: Natural history of cervical intraepithelial neoplasia: a critical review. In *Int J Gynecol Pathol* 12 (1993) 2, 186-192.
38. Papillo JL, Zarka MA, St John TL: Evaluation of the ThinPrep Pap test in clinical practice. A seven-month, 16,314-case experience in northern Vermont. In *Acta Cytol* 42 (1998) 1, 203-208.
39. Park IA, Lee SN, Chae SW, Park KH, Kim JW, Lee HP: Comparing the accuracy of ThinPrep Pap tests and conventional Papanicolaou smears on the basis of the histologic diagnosis: a clinical study of women with cervical abnormalities. In *Acta Cytol* 45 (2001) 4, 525-531.
40. Payne N, Chilcott J, McGoogan E Liquid-based cytology in cervical screening: a rapid and systematic review. 2000
41. Schiffman MH, Bauer HM, Hoover RN, Glass AG, Cadell DM, Rush BB, Scott DR, Sherman ME, Kurman RJ, Wacholder S, Stanton CK, Manos MM: Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. In *Journal of the National Cancer Institute* 85 (1993) 12, 958-964.

42. Schneider A, Hoyer H, Lotz B, Leistritz S, Kuhne-Heid R, Nindl I, Muller B, Haerting J, Durst M: Screening for high-grade cervical intra-epithelial neoplasia and cancer by testing for high-risk HPV, routine cytology or colposcopy. In *International Journal of Cancer* 89 (2000) 6, 529-534.
43. Sheets EE, Constantineee NM, Dinisco S, Dean B, Cibas ES: Colposcopically Directed Biopsies Provide a Basis for Comparing the Accuracy of ThinPrep and Papanicolaou smears. In *Journal of Gynecologic Techniques* 1 (1995) 27-33.
44. Sherman ME, Schiffman M, Herrero R, Kelly D, Bratti C, Mango LJ, Alfaro M, Hutchinson ML, Mena F, Hildesheim A, Morales J, Greenberg MD, Balmaceda I, Lorincz AT: Performance of a semiautomated Papanicolaou smear screening system: results of a population-based study conducted in Guanacaste, Costa Rica. In *Cancer* 84 (1998) 5, 273-280.
45. Syrjanen K, Kataja V, Yliskoski M: Natural history of cervical humanpapillomavirus lesions does not substantiate the biologic relevance of the Bethesda System. In *Obstet Gynecol* 79 (1992) 675-682.
46. Vassilakos P, Schwartz D, de Marval F, Yousfi L, Broquet G, Mathez-Loic F, Campana A, Major A: Biopsy-based comparison of liquid-based, thin-layer preparations to conventional Pap smears. In *J ReprodMed* 45 (2000) 1, 11-16.
47. Vessey MP, Villard-Mackintosh L: The epidemiology of hysterectomy: findings in a large cohort study. In *British Journal of Obstetrics and Gynaecology* 99 (1992) 402-407.
48. Weintraub J, Morabia A: Efficacy of a liquid-based thin layer method for cervical cancer screening in a population with a low incidence of cervical cancer. In *DiagnCytopathol* 22 (2000) 1, 52-59.
49. Wilbur D, Prey M, Miller W, Pawlick G, Colgan T: The AutoPap system for primary screening in cervical cytology. Comparing the results of a prospective, intended-use study with routine manual practice. In *Acta Cytol* 42 (1998) 1, 214-220.

Danksagung:

Für die Durchführung dieses Health Technology Assessments wurde ein interdisziplinäres Expertenpanel zur wissenschaftlichen Beratung der Autoren des vorliegenden Berichtes eingerichtet (siehe Abschnitt Expertenkontakte). Den Mitgliedern dieses Panels sei an dieser Stelle herzlich für Ihren wissenschaftlichen Rat und die Bereitstellung von Literatur, Dokumenten oder Daten gedankt. Frau Dr. Silke Siebert (Harvard Extension School, Boston) sei für die externe Überprüfung der in Text und Tabellen dargestellten Rechenergebnisse und die gründliche Durchsicht des Manuskripts gedankt. Für den Inhalt der vorliegenden Version einschließlich der darin enthaltenen Daten sind ausschließlich die Autoren dieses Berichtes verantwortlich.

Glossar

Begriff	Erklärung
Adequacy	Kategorie aus der Bethesda-Klassifikation, die eine Aussage zur Abstrichqualität liefert. Man unterscheidet dabei die Ausprägungen: Satisfactory (ausreichend gut beurteilbare Qualität), Satisfactory but limited by results (eingeschränkt beurteilbare Qualität) und Unsatisfactory (unzureichende Qualität).
Basisfallanalyse	Analyse eines entscheidungsanalytischen Modelles unter Verwendung bestimmter struktureller Annahmen und Ausprägungen für unsichere Parameter. Eine Variation dieser Parameter ist im Rahmen von > Sensitivitätsanalysen möglich, deren Ergebnisse dann mehr oder weniger vom Ergebnis der Basisfallanalyse abweichen.
Bethesda-Klassifikation	US-amerikanisches Klassifikationssystem zur Systematik der gynäkologischen Zytologie, wird bei den meisten internationalen Studien verwendet, keine Anwendung in der Routinezytologie in Deutschland.
Cut-off Werte	Schwellenwerte, die festgelegt werden, um bei kontinuierlichen Testgrößen eine Grenze zwischen 'test-positiv' und 'testnegativ' festzulegen.
Deflationierung	Rechnerische Bereinigung um die Wirkungen der Inflation auf Preise und Ausgaben.
Diskontierung	Rechenmethode, um Kosten und medizinischen Nutzen, die bei verschiedenen medizinischen Interventionen zu unterschiedlichen Zeitpunkten anfallen, vergleichbar zu machen. Die zukünftigen Werte für Kosten und medizinischen Nutzen werden durch Verwendung einer angemessenen > Diskontrate geringer gewichtet als Werte in der Gegenwart und in den entsprechenden Gegenwartswert (Barwert) umgerechnet.

Begriff	Erklärung
Diskontrate	<p>Rate, die verwendet wird, um aus zukünftigen Kosten und/oder Nutzwerten den Gegenwartswert (Barwert) zu ermitteln. Der Gegenwartswert berechnet sich folgendermaßen:</p> $GW = K / (1+r)^n;$ <p>GW = Gegenwartswert (Barwert), K = Kosten zum jeweiligen Zeitpunkt des Auftretens, r = jährliche Diskontrate, n = Laufzeit</p> <p>Bei der Wahl der Diskontrate kann man sich an verschiedenen Größen orientieren, z.B. der Rendite langfristiger Staatspapiere oder staatlicher Vorgaben. Meist werden die Werte 3% oder 5% verwendet.</p>
Dominanz	<p>Eine Strategie wird gegenüber einer Alternativstrategie als dominant bezeichnet, wenn sie sowohl effektiver als auch kostengünstiger als die Alternativstrategie ist.</p>
Dominanz, erweiterte	<p>Syn.: schwache Dominanz</p> <p>Eine Strategie wird im Rahmen von sich gegenseitig ausschließenden Strategien als erweitert dominiert (oder schwach dominiert) bezeichnet, wenn sie weniger effektiv und teurer ist als eine Linearkombination zweier anderer Strategien.</p>
Dysplasien	<p>Fehlbildung bzw. Fehlentwicklung von Zellen und Geweben (David; 1987).</p>
Effekt, inkrementeller	<p>Differenz zwischen dem Effekt einer Strategie und dem Effekt der Alternativstrategie.</p>
Entscheidungsanalyse	<p>Systematischer, expliziter und quantitativer Ansatz zur Entscheidungsfindung unter Unsicherheit. Die Wahrscheinlichkeiten der möglichen Ereignisse werden mit den Konsequenzen dieser Ereignisse dargestellt. Die Entscheidungsanalyse verwendet probabilistische Modelle (> Entscheidungsbaum, > Markov-Modell) zur Analyse von Primär- und/oder Sekundärdaten.</p>
Entscheidungsbaum	<p>Graphische Darstellung eines entscheidungsanalytischen Modelles. Enthält die zeitliche und logische Struktur des Entscheidungsproblem und alle relevanten alternativen Strategien, unsicheren Ereignisse, deren Eintrittswahrscheinlichkeiten sowie die zu erwartenden Ergebnisse.</p>

Begriff	Erklärung
Extremwertanalyse worst/best case analysis	Besondere Form der > Sensitivitätsanalyse. Alle unsicheren Einflussparameter werden so gewählt, dass sie sich maximal für bzw. gegen eine bestimmte Entscheidung auswirken. Damit kann geprüft werden, ob eine Entscheidung auch unter extremen Annahmen stabil bleibt.
Falsch-Negative Rate (FNR)	Anteil der falsch negativen Testergebnisse, bezogen auf alle Personen, die tatsächlich erkrankt sind.
Falsch-Positive Rate (FPR)	Anteil der falsch positiven Testergebnisse, bezogen auf alle Personen, die gesund sind.
Gesellschaftliche Perspektive	Umfassendster Blickwinkel für eine gesundheitsökonomische Evaluation, der alle Kosten und Gesundheitseffekte in der Gesellschaft berücksichtigt, unabhängig davon, für wen die Kosten bzw. Effekte anfallen.
Gesundheitszustand	Zustand der Gesundheit eines Individuums zu einem bestimmten Zeitpunkt. Zur einfacheren Durchführung von Entscheidungsanalysen wird meist eine überschaubare Zahl von diskreten Gesundheitszuständen verwendet, die dann mit Nutzwerten besetzt werden.
Goldstandard, diagnostischer	Der Goldstandard ist die bis anhin als beste und am zuverlässigsten eingestufte Methode zum Nachweis oder Ausschluss einer Erkrankung. Er kommt Idealerweise dem theoretischen Referenzstandard, der in der Lage ist, in Bezug auf die untersuchte Erkrankung mit vollständiger Sicherheit zwischen 'gesund' und 'krank' zu unterscheiden, am nächsten.
Inkrementelles Kosten-Effektivitäts-Verhältnis (IKEV)	> Kosten-Effektivitäts-Verhältnis.
Inkrementelles Kosten-Nutzwert-Verhältnis (IKNV)	> Kosten-Nutzwert-Verhältnis.
Inkrementelle Analyse	Im Gegensatz zur Analyse der durchschnittlichen Kosten und Effekte werden die im Vergleich mit der alternativen Strategie verursachten > zusätzlichen Kosten und > zusätzlichen Effekte einer medizinischen Intervention ins Verhältnis gesetzt.
Interobserver-Variabilität	Variabilität zwischen verschiedenen Untersuchern; wird z.B. bei dichotomen Variablen gemessen als Grad an Übereinstimmung (prozentuale Gesamtübereinstimmung) statistisch getestet mit der >Kappa-Statistik.

Begriff	Erklärung
Inzidenz	Die Inzidenz ist definiert als die Anzahl neuer Krankheitsfälle, die in einem bestimmten Zeitraum auftreten, bezogen auf die Bevölkerung mit gleichem Erkrankungsrisiko (Gordis; 2001).
Kappa-Statistik (k)	Sie bestimmt, inwieweit der Grad an Übereinstimmung ein rein zufälliges Maß überschreitet. (Gordis; 2001). Die gewichtete Kappa-Statistik misst die Divergenz (Nichtübereinstimmung, 'disagreement') von Diagnosen entsprechend der Bedeutung oder der Schwere des Irrtums. Eine perfekte Übereinstimmung ergibt ein $k = 1$, als gute Übereinstimmung gilt $k = 0.5$, bei rein zufälliger Übereinstimmung ist $k = 0$.
Kaufpreisparitäten	Austauschverhältnis zwischen zwei Währungen, berechnet mit dem Wert gewichteter Warenkörbe.
Kolposkopie	Lupenuntersuchung des Teils des Gebärmutterhalses, der in die Scheide hineinragt.
Konfidenzintervall	Wird auch als Vertrauensintervall oder Bereichsschätzer bezeichnet. Das 95%-Konfidenzintervall ist derjenige Bereich um einen erhobenen Punktschätzer (z.B. gemessene Sensitivität in einer Stichprobe), in dem der wahre Wert der Grundgesamtheit mit einer Wahrscheinlichkeit von 95% liegt.
Kosten, inkrementelle	Differenz zwischen den Kosten einer Strategie und den Kosten der Alternativstrategie.
Kosten, zusätzliche	> Kosten, inkrementelle.
Kosten-Effektivitäts-Verhältnis (KEV)	Der Quotient zwischen der Differenz der Kosten zweier Strategien und der Differenz der Effektivität dieser Strategien. Dabei wird die Effektivität in natürlichen Einheiten angegeben. Der Begriff wird auch häufig als Oberbegriff für > Kosten-Effektivitäts-Verhältnis (KEV) und > Kosten-Nutzwert-Verhältnis (KNV) verwendet.
Kosten-Nutzwert-Verhältnis (KEV)	Der Quotient zwischen der Differenz der Kosten zweier Strategien und der Differenz der Nutzwerte dieser Strategien. Dabei werden die Nutzwerte meist als qualitätskorrigierte Lebensjahre (QALYs) angegeben.
Kostenträger-Perspektive	Perspektive einer Institution, welche die Gesundheitsleistungen bezahlt.

Begriff	Erklärung
Likelihood Ratio (LR)	Im diagnostischen Kontext: der Quotient zwischen der bedingten Wahrscheinlichkeit für das Testergebnis bei Kranken und der bedingten Wahrscheinlichkeit für das Testergebnis bei Gesunden. Z.B. vergleicht die positive LR, wie viel mal häufiger bei der interessierenden Erkrankung ein richtig positiver Test (bei Kranken) als ein falsch positiver Test (bei Gesunden) vorkommt.
Markov-Annahme	Die Annahme, dass die > Übergangswahrscheinlichkeiten in > Markov-Modellen nicht von den in vorangegangenen Zyklen durchlaufenen > Gesundheitszuständen abhängen, sondern einzig und allein von den Gegebenheiten im aktuellen Zyklus ("Gedächtnislosigkeit bezüglich früherer Zyklen").
Markov-Modell	Mathematischer Modelltyp der Entscheidungsanalyse. Ein Markov-Modell enthält eine endliche Zahl von disjunkten und erschöpfenden > Gesundheitszuständen und besitzt Zeitintervalle, in denen die Übergangswahrscheinlichkeiten zwischen den Zuständen vom momentanen Zustand anhängen. Markov-Modelle sind anzuwenden, (i) wenn das Entscheidungsproblem Risiken im Zeitverlauf beinhaltet, (ii) wenn der Zeitpunkt des Auftretens eine Rolle spielt (time-to-event) oder (iii) wenn relevante Ereignisse mehr als einmal auftreten können. Man unterscheidet zwischen > Markov-Ketten und > Markov-Prozessen.
Metaanalyse	Eine Metaanalyse ist ein statistisches Verfahren, um die Effekte aus verschiedenen, aber vergleichbaren Studien quantitativ zusammenzufassen (Pooling). Das Ziel einer Metaanalyse ist es, durch Poolen der Daten deren statistische Aussagekraft zu erhöhen.
Nutzwert	Konzept in Ökonomie, Psychologie und Entscheidungsanalyse, welches sich auf die Präferenz bezüglich bestimmter Ergebniszustände bezieht. Im Kontext der Messung gesundheitsbezogener Lebensqualität reflektieren Nutzwerte die Präferenz für verschiedene > Gesundheitszustände auf einer gemeinsamen Skala.
Odds	Die Odds gibt die "Chance" an, mit der ein Ereignis eintritt. Der mathematische Zusammenhang der Odds mit der Wahrscheinlichkeit (P) lautet: Odds = $P/(1-P)$.

Begriff	Erklärung
Odds Ratio, diagnostische (DOR)	Beschreibt Chancenverhältnis eines positiven Tests bei einem Erkrankten im Verhältnis zu einem nicht Erkrankten.
Opportunitätskosten	Kosten einer Verwendung von Ressourcen bewertet mit dem monetären Nutzen der nächstbesten Alternative.
Perspektive	<p>Blickwinkel bzw. Sichtweise, aus der eine Entscheidungsanalyse bzw. eine gesundheitsökonomische Evaluation durchgeführt wird.</p> <p>Beispiele: Perspektive der Patienten, Leistungserbringer, Kostenträger, Hersteller von Arzneimitteln und medizinischen Geräten, Arbeitgeber, Gesundheitsbehörden, Gesellschaft. Die gesamtgesellschaftliche Perspektive ist die umfassendste.</p>
Präkanzerosen	Lang dauernde Gewebsveränderungen bzw. Erkrankungen, die, ohne selbst Krebs zu sein, mit hoher Wahrscheinlichkeit zu einem Krebs führen (David; 1987).
Prävalenz	Zahl der erkrankten Personen innerhalb der Bevölkerung zu einer bestimmten Zeit, geteilt durch die Anzahl der Gesamtbevölkerung zu dieser Zeit (Gordis; 2001).
Qualitätskorrigierte Lebenserwartung	<p>Engl.: quality-adjusted life expectancy</p> <p>Lebenserwartung, die über qualitätskorrigierte Lebensjahre (QALYs) anstelle nominaler Lebensjahre berechnet wurde.</p>
Rescreening	<p>Verfahren, bei dem Objektträger mit negativem Zytologiebefund im ersten Screening einer erneuten Musterung unterzogen werden. Formen:</p> <ul style="list-style-type: none"> - 10%-iges randomisiertes Rescreening - Gezieltes Rescreening bei Frauen mit Vorgeschichte positiver Befunde - Rapid Rescreening: zeitlich begrenztes Muster zytologischer Präparate mit einer Dauer von z.B. 30 – 90 sec
Schwellenwertanalyse	Typ der > Sensitivitätsanalyse, bei dem unsichere Einflussparameter über einen bestimmten Bereich variiert werden, um den Wert dieser Parameter zu bestimmen, für den sich die Entscheidung ändern würde.
Sensitivität	Die Sensitivität gibt die Wahrscheinlichkeit eines positiven Testbefundes bei erkrankten Personen an

Begriff	Erklärung
Sensitivitätsanalyse	<p>Mathematische Verfahren, welche die Auswirkung von Modellannahmen auf den Entscheidungsausgang untersuchen. Dabei werden strukturelle Annahmen, Wahrscheinlichkeiten und Zustandsbewertungen über einen gewissen Bereich variiert. Wichtige Ziele der Sa. sind die Prüfung der Stabilität der Entscheidung, die Identifizierung einflussreicher Parameter und die Ermittlung von Schwellenwerten.</p> <p>Man unterscheidet einerseits zwischen > deterministischer Sa. und > probabilistischer Sa. und andererseits zwischen Einweg- und Mehrweg-Sensitivitätsanalyse.; Sonderformen der Sa. sind die > Schwellenwertanalyse und die > Extremwertanalyse.</p>
Sensitivitätsanalyse, deterministische	> Sensitivitätsanalyse, bei der einzelne oder mehrere unsichere Einflussparameter jeweils über einen vorgegebenen Bereich variiert werden und die Ergebnisparameter in Abhängigkeit der Einflussparameter dargestellt werden.
Sensitivitätsanalyse, einfache	Sensitivitätsanalyse, bei der nur eine Annahme bzw. ein Parameter variiert wird.
Sensitivitätsanalyse, Einweg-	> Sensitivitätsanalyse, einfache.
Sensitivitätsanalyse, Mehrweg-	> Sensitivitätsanalyse, multiple.
Sensitivitätsanalyse, multiple	Sensitivitätsanalyse, bei der zwei oder mehr Annahmen bzw. Parameter simultan variiert werden.
Sensitivitätsanalyse, probabilistische	<p>Syn.: verteilungsorientierte Sensitivitätsanalyse.</p> <p>Entscheidungsanalytisches Simulationsverfahren, in welchem für jeden unsicheren Einflussparameter Werte zufällig und wiederholt aus den zugehörigen Verteilungen gezogen werden. Es resultiert eine Wahrscheinlichkeitsverteilung der erwarteten Ergebnisparameter.</p>
Spezifität	Die Spezifität gibt die Wahrscheinlichkeit eines negativen Testbefundes bei nicht erkrankten Personen an
Sterbetafelmethode	Mathematische Vorgehensweise, bei der die Mortalität oder die Morbidität einer geschlossenen Population in solch kleinen aufeinanderfolgenden Intervallen evaluiert wird, dass innerhalb des Intervalls von einer annähernd konstanten Mortalitäts- bzw. Inzidenzrate ausgegangen werden kann.

Begriff	Erklärung
Strategie	Eine festgelegte Sequenz von Entscheidungen, die über eine bestimmte Zeit zu treffen sind. Teil der Strategie sind feste Entscheidungsregeln, die sich der zu jedem Zeitpunkt gegebenen Informationen bedienen.
Übergangswahrscheinlichkeit	Die Wahrscheinlichkeit, in einem bestimmten Zeitintervall von einem bestimmten > Gesundheitszustand in einen anderen bestimmten Gesundheitszustand zu wechseln (s.a. > Markov-Modell).
Übergangswege	Mögliche Übergänge zwischen den verschiedenen > Gesundheitszuständen in einem > Markov-Modell.
Unabhängigkeit zweier Ereignisse	Zwei Ereignisse A und B nennt man unabhängig, wenn gilt: $P(A B) = P(A)$.
Utilitarismus	Theorie sozialer Gerechtigkeit, die darauf abzielt, einen möglichst großen durchschnittlichen Wohlstand zu erreichen. Viele der entscheidungsanalytischen Verfahren beruhen auf dem utilitaristischen Prinzip und wenden dieses für das Gut Gesundheit an.
Zahlungsbereitschaft	Bewertungstechnik, mit welcher der maximale Geldbetrag ermittelt wird, den ein Individuum zu zahlen bereit ist, um einen bestimmtes Gut oder eine Leistung (z.B. Erhöhung der Lebenserwartung, Erhöhung der Lebensqualität, Verminderung eines Morbiditätsrisikos) zu erlangen.
Zeithorizont	Zeitraum in einer Entscheidungsanalyse, über den medizinische Effekte und/oder Kosten gemessen werden.
Zykluslänge	Länge eines Zyklus innerhalb eines > Markov-Modelles.