

Die klinische Epidemiologie

in der ärztlichen Entscheidungsfindung

Habilitationsschrift
zur Erlangung der Lehrbefähigung
für das Fach

Klinische Epidemiologie

vorgelegt dem Fakultätsrat der Medizinischen Fakultät Charité
der Humboldt-Universität zu Berlin

von

Frau Dr. med. Regina Agnes Kunz

geboren am 3. Januar 1962 in Ansbach

Präsident: Prof. Dr. rer. nat. J. Mlynek

Dekan: Prof. Dr. Joachim W. Dudenhausen

Öffentlich-wissenschaftlicher Vortrag: 23. Oktober 2003

Gutachter:

1. Herr Univ.-Prof. Dr. med. Hans-Werner Hense
2. Herr Univ.-Prof. Dr. med. Henning Zeidler

Danksagung

Die in dieser Arbeit beschriebenen Projekte wären ohne die Anregungen vieler Kollegen und Freunde aus der nationalen und internationalen EbM-Szene nicht möglich gewesen. Ihnen möchte ich an dieser Stelle danken.

Mein besonderer Dank gilt Prof. D.L. Sackett und Prof. A. Oxman, Prof. H.-H. Neumayer, der diese Arbeit in den vergangenen Jahren mit wohlwollenden Ratschlägen kritisch begleitet hat sowie den Partnern und Mitarbeitern der verschiedenen Projekte: Dr. K.S. Khan, Dr. L. Fritsche, den Mitarbeitern der Kaiserin-Friedrich-Stiftung, insbesondere Prof. J. Hammerstein für die Unterstützung bei der Durchführung der EbM-Kurse, Prof. S. Müller-Lissner, Frau Dr. W. Zielinski, N.Rakowsky und den Mitarbeitern der Parkklinik Weissensee sowie meinen Freunden im Deutschen Netzwerk EbM, insbesondere Dr. G. Antes, Prof. H. Raspe, Prof. G. Ollenschläger und M. Lelgemann.

Die kontinuierliche Unterstützung und Begleitung von Frau Prof. G. Kaczmarczyk durch die Klippen der Habilitation möchte ich besonders hervorheben.

Inhaltsverzeichnis

Vorwort: Thema der Habilitationsschrift.....	7
Bearbeitete Fragestellungen.....	8
1 Entstehung der klinischen Epidemiologie und die Entwicklung von EbM.....	11
1.1 Die Entstehung der klinischen Epidemiologie.....	11
1.1.1 Entwicklung in Nordamerika	11
1.1.2 Großbritannien	12
1.1.3 Die (klinische) Epidemiologie in Deutschland.....	13
1.2 Inhalte der klinischen Epidemiologie.....	14
1.2.1 Die randomisierte kontrollierte Studie als Goldstandard des Wirksamkeitsnachweises.....	14
1.2.2 Meta-Analysen und systematische Übersichten von RCTs.....	15
1.2.3 Diagnostische Studien – die Herausforderung der laufenden Dekade.....	16
1.2.4 Weitere Fragestellungen und Entwicklungen.....	17
1.3 Von der klinischen Epidemiologie zur evidenzbasierten Medizin.....	18
1.3.1 Die Entwicklung zur evidenzbasierten Medizin.....	18
1.3.2 Was bedeutet „Evidenzbasierte Medizin“?	18
1.3.2.1 Die Definition.....	18
1.3.2.2 Der Evidenzpfad - Die Bedeutung der evidenzbasierten Medizin für das Gesundheitswesen	20
1.3.3 EbM in Deutschland.....	20
2 Beobachtung oder Experiment? Empirische Methodenforschung.....	22
2.1 Beobachtung oder Experiment – ein empirischer Vergleich (eigene Untersuchungen)23	
2.1.1 Das „Unpredictability Paradox“ – Randomisierte und nichtrandomisierte Studien im Vergleich.....	25
2.1.1.1 Fragestellungen.....	25
2.1.1.2 Methodik	25
2.1.1.3 Ergebnisse	26
2.1.1.4 Diskussion	34
2.1.1.5 Aktualisierung der Studie	35
2.2 Weiterentwicklung der Diskussion durch Folgestudien.....	36
2.2.1 Zusammenfassung der Folgestudien	36
2.2.1.1 Kritische Analyse der gegenwärtigen Datenlage	37
2.2.1.2 Variabilität zwischen Studiendesigns: wünschenswert oder problematisch ?.....	39
2.2.2 Forschungsbedarf: Heterogenität - klinisch und methodisch.....	39
2.2.3 Beobachtungsstudien – eine Standortbestimmung	40
2.3 Datenbank für methodische Studien – Grundstock für eine empirische Methodikforschung (eigene Untersuchungen).....	41
2.3.1 Fragestellung.....	41
2.3.2 Methodik.....	42
2.3.2.1 Eingesetzte Datenbanken.....	42
2.3.2.2 Einschlusskriterien.....	42
2.3.2.3 Entwicklung der Suchstrategien	43

2.3.3	Ergebnisse	45
2.3.3.1	Ausbeute der MEDLINE-Suche	45
2.3.3.2	Ausbeute der Suche im Science Citation Index	47
2.3.3.3	Sekundäre Fragestellungen	48
2.3.4	Diskussion	49

3 Risiko und Wirksamkeit, Behandlungsschwelle und Bias: Potential für klinische

Fehlentscheidungen..... 51

3.1 Die klinische Epidemiologie in der ärztlichen Entscheidungsfindung 52

3.1.1	Wirksamkeit von Interventionen	52
3.1.1.1	Relative Maßzahlen	52
3.1.1.2	Absolute Maßzahlen	53
3.1.1.3	Übertragbarkeit auf den Versorgungsalltag: Grenzen und Beschränkungen der Maßzahlen	54
3.1.2	Nutzen für den Patienten: Ausgangsrisiko und Risikoreduktion	55
3.1.2.1	Wahrnehmung von Ausgangsrisiko und Behandlungsnutzen – Drei Beispiele	56
3.1.2.2	Konsequenzen für die Versorgung.....	57
3.1.3	Behandlungsschwellen und ihre Determinanten	58
3.1.3.1	Wo liegen die Behandlungsschwellen? Eine empirische Momentaufnahme	58
3.1.3.2	Die Behandlungsschwelle beim individuellen Patienten – Grenzlinie zwischen Nutzen, Schaden und persönlichen Werten	59

3.2 Behandlungsschwelle und Bias: Potential für klinische Fehlentscheidungen (eigene

Untersuchungen)..... 59

3.2.1	Fragestellung	60
3.2.2	Methodik	60
3.2.2.1	Klinischer Rahmen des Modells	60
3.2.2.2	Methodische Grundbegriffe des Modells.....	61
3.2.3	Ergebnisse	62
3.2.3.1	Individualisierung der Behandlung.....	62
3.2.3.2	Fehlende Präzision bei der Schätzung des Behandlungseffekts.....	63
3.2.3.3	Die Auswirkung von Bias in Forschungsergebnissen auf ärztliche Entscheidungen	64
3.2.4	Schlussfolgerungen	67
3.2.4.1	Konsequenzen für den behandelnden Arzt	67
3.2.4.2	Bias aus Beobachtungsstudien.....	68
3.2.4.3	Implikationen für die Praxis	69

3.3 Der Leitlinien-Clearingbericht Hypertonie (Eigene Ergebnisse) 69

3.3.1	Das Leitlinien – Clearingverfahren der Ärztlichen Zentralstelle Qualitätssicherung	69
3.3.2	Zielsetzung des Leitlinien-Clearingverfahrens Hypertonie	70
3.3.2.1	Methode.....	70
3.3.2.2	Ergebnisse	71
3.3.3	Schlussfolgerungen	74

4	Evidenzbasierte Medizin – Neue Fertigkeiten?.....	75
4.1	EbM – ein Modell für lebenslanges Lernen in der Medizin.....	76
4.1.1	EbM-Kurse nach dem McMaster-Format (Didaktik der EbM).....	77
4.1.2	EbM in der Weiterbildung - Ausländische Erfahrungen.....	78
4.1.2.1	„EbM-Practitioners“: Modell der ärztlichen Weiterbildung an der McMaster-University.....	78
4.1.2.2	EbM in US-amerikanischen Residency-Programmen.....	79
4.1.3	EbM in Deutschland.....	80
4.1.3.1	Qualifizierungsmöglichkeiten in EbM in Deutschland.....	80
4.1.3.2	EbM in der institutionalisierten Aus- und Weiterbildung.....	80
4.1.4	Was bewirken EbM-Kurse?.....	81
4.1.4.1	Ernüchternde Datenlage.....	81
4.1.4.2	Komplexität der Intervention – Komplexität des Nachweises.....	82
4.1.4.3	Anforderungen an Evaluationsinstrumente.....	82
4.1.4.4	Wann wird Evaluation zur Forschung?.....	83
4.2	Die Berliner EbM-Studie – Nutzwert von EbM-Kursen (eigene Untersuchungen)	84
4.2.1	Der Berliner Gegenstandskatalog.....	84
4.2.1.1	Zielsetzung.....	84
4.2.1.2	Methode der Erstellung.....	85
4.2.1.3	Ergebnisse.....	85
4.2.1.4	Diskussion.....	86
4.2.1.5	Schlussfolgerungen.....	86
4.2.1.6	Das „Curriculum Evidenzbasierte Medizin“.....	87
4.2.2	Der „Berliner Fragebogen“: Entwicklung und Validierung eines Instruments zur Messung von EbM-Kenntnissen	87
4.2.2.1	Methodik.....	88
4.2.2.2	Ergebnisse.....	89
4.2.3	Die Berliner EbM-Studie.....	92
4.2.4	Untersuchung der Wirksamkeit von EbM-Training am Beispiel des Berliner Grundkurses EbM. Eine kontrollierte Vorher-Nachher („Before-After“)-Studie (eigene Untersuchungen).....	92
4.2.4.1	Methodik.....	92
4.2.4.2	Ergebnisse.....	93
4.2.4.3	Diskussion.....	95
5	EbM - Praxisbeispiele und Modellprojekte.....	98
5.1	Evidenzbasierte Medizin - eine Herausforderung für praktizierende Ärzte?.....	99
5.1.1	Praxisprojekte zur Implementierung von EbM in die Versorgung.....	100
5.1.1.1	Prozessorientierte EbM-Projekte: FRONTLINE-Projekt und PACE-Projekt.....	101
5.1.1.2	EbM-Versorgungsstudien.....	102
5.1.1.3	Schlussfolgerungen.....	103

5.2	Das Park-Klinik Weissensee - EbM-Projekt (eigene Untersuchungen).....	104
5.2.1	Das Gesamtprojekt.....	104
5.2.2	EbM-gestützte Handlungsempfehlungen in Arztbriefen.....	106
5.2.2.1	Hintergrund.....	106
5.2.2.2	Fragestellung der Studie.....	107
5.2.2.3	Studiendesign und Methodik.....	107
5.2.2.4	Qualität der Arztbriefe - ein Audit.....	114
5.2.2.5	Ergebnisse.....	114
5.2.2.6	Diskussion.....	120
5.3	EbM-Informationsservice: Ein Pilotprojekt mit Rechercheangebot und kritischer Literaturbewertung (eigene Untersuchungen).....	123
5.3.1	Fragestellung.....	123
5.3.2	Methodik.....	123
5.3.3	Ergebnisse.....	124
5.3.4	Schlussfolgerungen.....	125
5.4	Implementierung in Deutschland.....	126
5.4.1	Praxisprojekte.....	126
5.4.2	Voraussetzungen für eine erfolgreiche Implementierung.....	127
6	Zusammenfassung und Perspektive.....	130
	Literaturverzeichnis.....	136
	Anhang.....	150

Vorwort: Thema der Habilitationsschrift

Mit der gesetzlichen Verankerung von evidenzbasierten Leitlinien und Disease-Management-Programmen hat das Konzept der evidenzbasierten Medizin (EbM) zur Lösung gesundheitspolitischer Probleme innerhalb von wenigen Jahren Einzug in das deutsche Gesundheitswesen gehalten. Die Aufnahme und Verbreitung von EbM unter den praktizierenden Ärzten und ihre Integration in die Patientenversorgung fiel wesentlich zurückhaltender aus. Dafür gibt es eine Vielzahl von Gründen: Unkenntnis und Unverständnis der Methodik der klinischen Epidemiologie, Gefühl der Bedrohung der professionellen Autonomie, Aufkündigung der Tradition physiologischer Begründungen und Schulendenkens. Das Konzept hinterfragt die Wurzeln der bisher praktizierten Medizin und die Qualität der bisher geleisteten Arbeit.

Auf diesem schwierigen Terrain der evidenzbasierten Patientenversorgung bewegt sich die vorliegende Habilitationsschrift mit dem Thema: *„Die klinische Epidemiologie in der ärztlichen Entscheidungsfindung“*. Der Titel beschreibt zwei Pole, die scheinbar nur schwer vereinbar sind: gruppenbezogener Erkenntnisgewinn aus der patientenorientierten Forschung und die Versorgung individueller Patienten in ihren persönlichen Lebensumständen. Die evidenzbasierte Medizin möchte diese beiden Pole verbinden. Der EbM-Pfad mit seinen fünf Schritten (s. Kapitel 1) beschreibt die einzelnen Stationen: Evidenzproduktion, Evidenzsynthese, evidenzbasierte Gesundheitsstrategien („Policies“), Umsetzung der Gesundheitsstrategien und evidenzbasierte Patientenbehandlung.

Evidenzbasierte Medizin bedeutet Transparenz, eine kritische Einstellung auch gegenüber den eigenen Prämissen und den Anspruch, diese mit empirisch erhobenen Daten („evidence“) zu belegen. Die Überprüfung von logischen Argumenten der EbM und ihres Anspruchs, eine hochwertige Patientenversorgung zu gewährleisten, zieht sich als Leitmotiv durch die Habilitationsschrift. Zu nahezu jedem Schritt des EbM-Pfads wurden eine oder mehrere Studien durchgeführt. Mit der Untersuchung fokussierter Fragestellungen leisten die einzelnen Projekte einen Beitrag zu laufenden nationalen und internationalen Diskussionen um die empirische Bewertung der evidenzbasierten Medizin und ihrer Rolle in der Gesundheitsversorgung.

Bearbeitete Fragestellungen

Die **Einleitung (Kapitel 1)** ist ein kurzer geschichtlicher Abriss über die Entstehung der klinischen Epidemiologie aus der klassischen Epidemiologie und ihre Weiterentwicklung zur „evidenzbasierten Medizin“, die durch ihre spezifischen Fragestellungen und Schwerpunkte zu einer eigenen Entität geworden ist. Dabei wird auch auf die besonderen Umstände in Deutschland eingegangen, wo das Gesundheitssystem durch das Fehlen einer klinisch-epidemiologischen Basis von den Anliegen der EbM-Bewegung unvorbereitet getroffen wurde.

Die erste Studie „**Beobachtung oder Experiment**“ (2.1) hinterfragt eine zentrale These der klinischen Epidemiologie: „Lässt sich die Bedeutung, die der Randomisierung in klinischen Studien für eine unverzerrte Effektmessung eingeräumt wird, durch Vergleiche mit nichtrandomisierten Studien belegen?“ Durch den empirischen Ansatz erhielt diese seit mehr als 30 Jahren laufende Diskussion einen neuen Impuls, der in den darauffolgenden Jahren von verschiedenen Arbeitsgruppen aufgegriffen und durch weitere, im N Engl J Med und in JAMA publizierte methodische Studien weiterentwickelt wurde (2.2). Im Rahmen unserer Untersuchung wurde der Grundstock für ein Register für empirische Methodikstudien gelegt (**Database of Trials of Reviews and Trials (TORTS)**), aus dem später die Cochrane Review Methodology Database (CRMD) hervorging (2.3). Die in diesem Kontext durchgeführten Untersuchungen werden ebenfalls dargestellt.

Doch was haben diese für Methodenwissenschaftler spannenden Überlegungen zu Bias mit der täglichen ärztlichen Entscheidungsfindung zu tun? Die Simulationsstudie „**Behandlungsschwelle und Bias**“ (3.2) untersucht für unterschiedliche Risikokonstellationen und auf der Grundlage realer Patientendaten, wie verzerrte Studienergebnisse zu klinischen Fehlentscheidungen führen können. Diese Simulation beruht auf grundlegenden klinisch-epidemiologischen Konzepten zur Effekt- und Risikobeschreibung und der Wirksamkeitsbewertung, die im klinischen Alltag routinemäßig angewandt werden. Die Einführung zu dieser Studie (3.1) reflektiert anhand von aktuellen Untersuchungen die (unzureichenden) Kompetenzen von Ärzten im Umgang mit diesen Konzepten und unterstreicht damit die Bedeutung der Simulationsstudie im klinischen Kontext. Diese Punkte werden in dem nächsten Projekt, dem **Leitlinien-Clearingverfahren** der Ärztlichen Zentralstelle

Qualitätssicherung für das Krankheitsbild **Hypertonie (3.3)** aufgegriffen, bei dem ich als Expertin beteiligt war. In dem Clearingverfahren, das wesentlich von den EbM-Prinzipien Transparenz und Begründbarkeit getragen wird, wurden nach einem internationalen Leitlinienvergleich Empfehlungen für eine nationale Hypertonieleitlinie abgegeben, in der den klinisch-epidemiologischen Konzepten „Risiko und Wirksamkeit beim Management von Hypertoniepatienten“ mehr Bedeutung beigemessen werden soll.

Workshops zur Vermittlung von Grundkenntnissen in EbM werden inzwischen überall in Deutschland angeboten. Die aufwändigen EbM-Kurse haben jedoch nur einen Sinn, wenn Wissen und Fertigkeiten in EbM wirklich verbessert werden. Da dies international wiederholt in Frage gestellt wurde, führten wir in unseren eigenen Berliner EbM-Kursen eine mehrjährige Studie durch (**Kapitel 4**): Auf der Grundlage definierter Lerninhalte unseres EbM-Kurses („**Berliner Gegenstandskatalog EbM**“ 4.2.1) entwickelten und validierten wir ein entsprechendes Instrument („**Berliner Fragebogen**“ 4.2.2), mit dem in einer prospektiven Studie der Zuwachs an Wissen und Fertigkeiten untersucht wurde („**Berliner EbM-Studie**“ 4.2.3).

Doch lässt sich eine evidenzbasierte Patientenversorgung in der Praxis wirklich umsetzen? Dieser Frage widmet sich das letzte Kapitel. Nach einer kurzen Einführung über internationale EbM-Projekte und EbM-Versorgungsstudien wird das Projekt einer Berliner Klinik vorgestellt, evidenzbasierte Arbeitsprinzipien systematisch und klinikweit einzuführen. Innerhalb des Projekts untersuchten wir in einer clusterrandomisierten Studie an der Schnittstelle stationäre / ambulante Versorgung, ob man mit kurzen evidenzbasierten Erläuterungen zu neu angesetzten Behandlungen die Hausärzte motivieren kann, die Behandlung fortzusetzen („**Evidenzbasierte Handlungsempfehlungen in Arztbriefen**“ 5.2.2). Ein zweites Praxisprojekt pilotiert den Aufbau eines **EbM-Informationsservices (5.3)** an einem Universitätsklinikum. In **Kapitel 6** werden die Ergebnisse der einzelnen Studien zu EbM zusammengefasst und es wird versucht, eine Perspektive für eine evidenzbasierte Patientenversorgung und die dazu notwendigen Voraussetzungen zu entwickeln.

Für die vorliegende Habilitationsschrift wurden die eigenen Publikationen teilweise übersetzt, mit weiterem nicht publiziertem Material ergänzt und in die diskutierten Themen integriert. Jedes Kapitel beginnt mit einer Zusammenfassung des Kapitelinhalts, in der die eigenen Arbeiten von den Studien der Kontextdiskussion abgegrenzt werden und einer Auflistung der dem Kapitel zugrundeliegenden eigenen Publikationen.

Aus Gründen der Lesbarkeit wurde bei der Bezeichnung von Personen im Allgemeinen nur die männliche Sprachform verwendet. Die Aussagen beziehen sich auf Männer und Frauen in gleicher Weise.

1 Entstehung der klinischen Epidemiologie und die Entwicklung von EbM

1.1 Die Entstehung der klinischen Epidemiologie

1.1.1 Entwicklung in Nordamerika

Der Begriff „Klinische Epidemiologie“ geht zurück auf den Infektiologen John R. Paul, Yale-University (1893–1971), der für die Erforschung von (Infektions-) Krankheiten einen neuen, empirischen Ansatz forderte, bei dem die in individuellen Patienten beobachteten Phänomene vor einer Verallgemeinerung in größeren Gruppen bestätigt werden müssten. Diese „... basic new science for preventive medicine...“¹ bezeichnete Paul als „clinical epidemiology“. Mit dem kürzlich verstorbenen Alvan R. Feinstein (1926–2001), einem Mathematiker, Internisten und klinischen Epidemiologen, der ebenfalls an der Yale University tätig war, fand er einen enthusiastischen Nachfolger seiner Ideen. Wie für Paul war auch für Feinstein die sorgfältige klinische Beobachtung Ausgangspunkt der Forschung in größeren Kollektiven. Wiewohl er mit seinen Ideen andere Querdenker wie D.L. Sackett zur Entwicklung neuer Konzepte und Methoden inspirierte, blieb er mit seinem Konzept der klinischen Epidemiologie („... *the clinicostatistical study of diseased populations...*“²) stark der beobachtenden Epidemiologie verhaftet. Erst spät integrierte er auch die Patientenversorgung in sein Verständnis, als er 19985 schrieb: „*Clinical epidemiology represents the way in which classical epidemiology,, has been enlarged to include clinical decisions in the personal encounter care for individual patients*“³).

Im Jahre 1967 wurde an der Medizinischen Fakultät der McMaster-University, einer Reformfakultät in Kanada, von David L. Sackett, einem Internisten, Nephrologen und Epidemiologen, die Abteilung „Clinical Epidemiology and Biostatistics“ gegründet. Damit erhielt Sackett, der damals in engem Austausch mit A.L. Feinstein stand, die Gelegenheit, seine Ideen und Vorstellungen von „klinischer“ Epidemiologie („... *the application of epidemiology and biometric methods to the study of diagnosis and therapy by a clinician who provides direct patient care*“⁴) umzusetzen, die sehr stark in der Patientenversorgung verwurzelt waren und bei denen der klinische Epidemiologe eine Doppelrolle als „Arzt und klinischer Forscher“ innehatte. Inhaltlich lag der Fokus auf der Erforschung der beiden

Kernbereiche klinischer Tätigkeit, Therapie und Diagnose, um Ärzten robuste und für die tägliche Praxis relevante Informationen zur Verfügung zu stellen.

Durch die dynamische Einbindung der Methoden benachbarter Fächer (Biometrie, Bevölkerungsepidemiologie, Sozialwissenschaften, Gesundheitsökonomie oder Psychologie) entwickelte die klinische Epidemiologie ihr eigenes Profil und konnte für neue dringende Probleme, wie den Fragen nach der Qualität oder Angemessenheit der Versorgung, umsetzbare Antworten anbieten. Damit wurde sie für die Entwicklung von Behandlungsstrategien in der Patientenversorgung wie auch von Steuerungsstrategien im Gesundheitswesen⁵ zu einer zunehmend wichtigen, weil robusten Informationsquelle und gewann den maßgeblichen Einfluss auf gesundheitspolitische Entscheidungen, den wir heute sehen.

1.1.2 Großbritannien

Unabhängig davon fand in Großbritannien unter dem schottischen Epidemiologen A.L. Cochrane (1909–1988) eine parallele Entwicklung statt. Im Gegensatz zu seinen Zeitgenossen lag für Cochrane die Grundlage des Wirksamkeitsnachweises („*Effectiveness*“) in empirischen Vergleichen von Behandlungsalternativen, vorzugsweise von randomisierten Studien⁶ und er forderte deshalb mit seiner berühmten Anklage „*It is surely a great criticism of our profession that we have not organised a critical summaryof all relevant randomised controlled trials*“⁷ die Erstellung eines systematischen Registers aller verfügbaren Studien.

Anfang der 80er Jahren wurde diese Herausforderung von Iain Chalmers, einem englischen Gynäkologen aufgegriffen. Mit dem erfolgreichen Aufbau einer Datenbank über alle randomisierten Studien der Perinatalmedizin⁸ und der anschließenden Zusammenfassung in Meta-Analysen^{9; 10} demonstrierte er die Umsetzbarkeit von A. Cochranes Vision. 1992 wurde in Oxford das Cochrane Centre gegründet, ein Jahr später die Internationale Cochrane Collaboration auf den Weg gebracht, mit dem Ziel, nach Chalmers' Vorbild über eine weltweite Zusammenarbeit das Register auf alle Gesundheitsbereiche auszudehnen. Innerhalb von acht Jahren entstanden weltweit 15 Cochrane Zentren, das Deutsche Cochrane Zentrum in Freiburg wurde 1997 gegründet. Heute umfasst die Collaboration mehr als 50 Reviewgruppen und mehr als 5 000 Mitarbeiter. Ihr Produkt, die Cochrane

Library, enthält mehr als 2 400 Reviews und Protokolle sowie ein weltweit einmaliges Controlled Trials Register mit über 350 000 Studien.

1.1.3 Die (klinische) Epidemiologie in Deutschland

Die internationalen Entwicklungen hin zu einer (klinischen) Epidemiologie mit klinischer patientenorientierter Forschung sind an Deutschland weitgehend vorbeigegangen, obwohl in Deutschland zeitgleich mit dem Infektiologen J.R. Paul in Yale kritische Ärzte wie Paul Martini aktiv waren¹¹. Paul Martini verfolgte ähnliche Vorstellungen und methodische Überlegungen und entwarf in seiner „Methodenlehre der Therapeutischen Untersuchung“¹² Studienkonzepte, die heute noch modern anmuten, die aber in der englischsprachigen Welt nicht wahrgenommen wurden¹¹. Die deutsche Sozialmedizin, die zu Beginn des Jahrhunderts über eine internationale Reputation verfügte, hatte im Dritten Reich durch den Exodus jüdischer Ärzte¹³ und durch die Entwicklung zur Rassenhygiene und deren aktiver Umsetzung¹⁴ schwere Einbußen erlebt, von deren Folgen sie sich auch nach dem 2. Weltkrieg für lange Zeit nicht erholte. Nach dem 2. Weltkrieg wurde der Bereich „öffentliche Gesundheit“ und seine Forschungsstrukturen beim Wiederaufbau des Gesundheitswesens weitgehend vernachlässigt. Als Ende der 70er Jahre die Defizite an fundierten Daten für gesundheitspolitische Entscheidungen offensichtlich wurden, begann man mit dem Aufbau eines multidisziplinären „Public Health“-Bereichs nach angelsächsischem Vorbild¹⁵, jedoch mit einer stark soziologischen Ausrichtung. Die klassische Epidemiologie mit ihrem quantitativen Ansatz fand durch das Engagement einzelner Epidemiologen wenigstens im Rahmen von (international besetzten) Sommerschulen (Münster oder Ulm) ein Forum. Mit dem Sonderprogramm des Deutschen Akademischen Austauschdienstes für Epidemiologie wurde Anfang der 90er Jahre eine neue Initiative gestartet, um im Ausland eine kritische Masse an Ärzten zu Epidemiologen auszubilden. Dennoch blieb die universitäre Epidemiologie in Deutschland im Wesentlichen in der Biometrie angesiedelt, ein ärztlicher Zugang im Sinne von Paul, Feinstein oder Sackett wurde kaum erschlossen. Mit der gegenwärtigen Förderung von „Koordinierungszentren Klinischer Studien“ (KKS) durch das BMBF wird der strukturelle Aufbau und das Know-how zur Durchführung klinischer Studien substanziell weiterentwickelt. Postgraduiertenausbildungsgänge in Epidemiologie mit international anerkanntem

Masterabschluss werden aufgebaut (z.B. Fakultät für Gesundheitswissenschaften, Bielefeld). Dennoch sind die Ausbildungsmöglichkeiten in klinischer Forschung in Deutschland unverändert unterentwickelt, wenn man vergleicht, dass an 15 der 16 der kanadischen Medical Schools ein Postgraduiertenstudiengang in klinischen Forschungsmethoden („Clinical Research Methods“; Sackett, persönliche Kommunikation) angeboten wird.

1.2 Inhalte der klinischen Epidemiologie

Die Kernthemen der klinischen Epidemiologie beschäftigen sich primär mit Fragen aus der Patientenversorgung: Therapie und Prävention, Diagnostik und Screening, Prognose, Nebenwirkungen und Schaden, sekundär kamen Fragen über Angemessenheit oder Qualität der klinischen Versorgung hinzu. Obwohl durch eine systematische Forschung alle Bereiche besser erschlossen wurden, haben die wesentlichen methodischen Entwicklungen im Bereich Therapie / Prävention stattgefunden.

1.2.1 Die randomisierte kontrollierte Studie als Goldstandard des Wirksamkeitsnachweises

In der klinischen Forschung waren durch die Entwicklung und Etablierung der randomisierten Studie (RCTs) als das am besten geeignete Design für einen glaubwürdigen Wirksamkeitsnachweis von Interventionen entscheidende Impulse gesetzt worden, dem allerdings heftige Debatten zwischen den „Trialists“ und den „Observationalists“ vorangegangen waren. Während die „Trialists“ (D.L. Sackett¹⁶, T.C. Chalmers¹⁷, H. Sacks¹⁸) argumentierten, dass nur durch die Randomisierung eine gleichmäßige Verteilung der bekannten und, noch wichtiger, der unbekanntem Einflussgrößen sichergestellt werden könne, vertraten die „Observationalists“ (u.a. A.R. Feinstein und R.I. Horwitz¹⁹) die Ansicht, dass dieses Ziel auch durch sorgfältige Planung und Durchführung von Beobachtungsstudien erreichbar sei. Letztendlich waren es die Trialists, die mit dem wiederholten unerwarteten Nachweis erhöhter Mortalität oder Morbidität von etablierten, pathophysiologisch gut begründeten Behandlungsmethoden¹⁹⁻²⁵ in RCTs den Diskurs zugunsten des experimentellen Designs entschieden.

Auf neue Fragestellungen und Anforderungen folgten Weiterentwicklungen des randomisierten Designs: Die „large simple trials“ für große Populationen bei häufigen Erkrankungen²⁶⁻³⁰, der N-of-1 Trial als eine individualisierte randomisierte Studie^{31; 32}, die Clusterrandomisierung bei Interventionen in interagierenden Populationen³³⁻³⁵ (s.a. Kapitel 5). Statt der Bewertung der Wirksamkeit unter Idealbedingungen („*Efficacy*“-Studie) rückte die Gesamtbewertung inklusive schlechter Compliance oder Studienabbruch durch Nebenwirkungen^{36; 37} in den Mittelpunkt der Betrachtung („*Effectiveness*“-Studie). Bei steigendem Informationsbedarf über die Kosten „pro Einheit Nutzen“ wurden gesundheitsökonomische Aspekte integriert („*Efficiency*“-Studien³⁸). Auch der Gegenstand der klinischen Forschung wandelte sich: Neben medikamentösen Behandlungen wurden und werden zunehmend auch nichtmedikamentöse Interventionen wie chirurgische Eingriffe^{39; 40} oder die Wirksamkeit ganzer Versorgungseinheiten⁴¹ auf den Prüfstand gestellt.

Durch seine Fokussierung auf den Wirksamkeitsnachweis ermöglicht der RCT auch eine faire Prüfung alternativer Heilverfahren wie der Homöopathie⁴²⁻⁴⁴, der Akupunktur⁴⁵⁻⁴⁷, der Phytotherapie⁴⁸⁻⁵⁰ oder der Physiotherapie^{51; 52}. Allerdings sehen darin nur wenige Vertreter unkonventioneller Heilmethoden^{53; 54} eine ernsthafte Chance, die Wirksamkeit ihrer Methoden mit wissenschaftlich anerkannten Methoden unter Beweis zu stellen.

Mit der veränderten Perspektive - weg von der Ätiologie hin zur Frage nach der Wirksamkeit - findet in Forschung und Versorgung eine fundamentale Verschiebung der medizinischen Argumentations- und Beweisführung statt, die von den Gründern der „Evidence-based Medicine“ als Paradigmenwechsel⁵⁵ bezeichnet wurde. Es gibt Kritiker, die in dieser Entwicklung eine ernsthafte Bedrohung der Grundlagenforschung der modernen Medizin befürchten. Angesichts des enormen Aufholbedarfs der klinisch-evaluativen Forschung erscheint diese Sorge für absehbare Zeit unbegründet.

1.2.2 Meta-Analysen und systematische Übersichten von RCTs

In dem Maße, wie die „Oxford Database of Perinatal Trials“ an Umfang zunahm (s.o.) wurde auch die Notwendigkeit für eine weitere Verdichtung der Informationen immer offensichtlicher. Neue Impulse kamen aus den Sozialwissenschaften, wo mit der „Meta-Analyse“ eine innovative Methode entwickelt

worden war⁵⁶. Diese Technik wurde übernommen, auf den medizinischen Bereich adaptiert und bei der Zusammenfassung der Studien der Datenbank eingesetzt. Die Meta-Analyse begann, trotz erheblichem initialen Widerstand⁵⁷⁻⁶¹ sich auch in der Medizin zu etablieren, was durch die rasche und selbstkritische Entwicklung einer robusten Methodologie deutlich begünstigt wurde. Neue Biasformen wurden aufgedeckt, die die öffentliche Wahrnehmung der Wirksamkeit von Maßnahmen verzerrten: Publikationsbias, die ausschließliche Publikation von Studien mit positivem Ergebnis^{28; 62-67}, und „Language Bias“, die bevorzugte Publikation von positiven Studienergebnissen in englischsprachigen Zeitschriften^{68; 69}. Techniken wie der „Funnel Plot“ wurden zum Nachweis fehlender (negativer) Studien in Meta-Analysen entwickelt^{70; 71} und die weltweite Erstellung prospektiver Studienregister eingefordert^{72; 73}. Wissenschaftliches Fehlverhalten wie nicht gekennzeichnete Doppel-, Dreifach-, Vierfachpublikationen derselben Ergebnisse wurde dadurch aufgedeckt⁷⁴.

Bei der systematischen Aufarbeitung der Primärstudien zeigten sich erhebliche Qualitätsschwankungen, was die Diskussion um die methodische Qualität von Primärstudien erneut anregte: Welchen Schutz bieten einzelne methodische Verfahren für eine unverzerrte Endpunktschätzung? Scores zur Bewertung der Studienqualität wurden erarbeitet⁷⁵ und ihre Validität durch empirische Überprüfung wieder in Frage gestellt⁷⁶. Die Ergebnisse der Reviewtätigkeiten hatten auch Folgen für das Publikationswesen. Als Reaktion auf oft unvollständige, nicht selten auch irreführende Abstracts und Langfassungen der Studienberichte wurden von Methodikern und Herausgebern führender Journals Kriterien zur Standardisierung von informativen Abstracts und ausführlichen Studienberichten (CONSORT-Statement^{77; 78}, QUORUM-Statement⁷⁹, MOOSE-Statement⁸⁰) entwickelt, die von vielen internationalen und nationalen Zeitschriften übernommen wurden. Damit wird mittel- und langfristig ein nicht zu unterschätzender Beitrag für die Durchführung hochwertiger Studien geleistet.

1.2.3 Diagnostische Studien – die Herausforderung der laufenden Dekade

Neue Technologien in der Bildgebung, hochspezifische Labortests, Nachweis genetischer Marker bedeuten einen Quantensprung bei den diagnostischen Möglichkeiten. Zum Problem werden die neuen Technologien durch ihre rasche unkritische Übernahme in die Routineversorgung, ohne ausreichende

Klärung der Testcharakteristika in den unterschiedlichen Settings. Undiskutiert bleiben die Frage des Mehrwerts gegenüber etablierten oder alternativen Testverfahren oder die Frage nach ihrer Bedeutung in der Kette der Entscheidungsfindung. Angesichts der Zunahme an neuen Testverfahren, des in der Versorgung zu beobachtenden „diagnostischen Overkills“ und der daraus resultierenden finanziellen Belastung der Gesundheitssysteme besteht hier dringender Forschungs- und Diskussionsbedarf.

Trotz gut entwickelter methodischer Konzepte zur Evaluation diagnostischer Verfahren⁸¹ demonstrieren wiederholte Längsschnittuntersuchungen erhebliche Defizite⁸²⁻⁸⁴, die sich, anders als bei den RCTs, in den vergangenen 20 Jahren nur marginal gebessert haben^{84; 85}.

In den letzten Jahren zeichnet sich jedoch in der klinischen Epidemiologie die Entwicklung einer neuen Forschungsagenda ab: Durch die Thematisierung der erheblichen methodischen Defizite diagnostischer Studien bei der Erstellung systematischer Übersichten steigt das Qualitätsbewusstsein⁸⁴. Neue Konzepte für die Erstellung von Referenzstandards bei inkomplettem Goldstandard werden gesucht. Methoden zur weiteren Untersuchung heterogener Studienergebnisse jenseits von Qualitätsunterschieden^{86; 87} müssen entwickelt werden. Parallel zu den Interventionsstudien wurde auch bei den diagnostischen Tests mit einer empirischen Methodikforschung begonnen⁸⁶. Insgesamt steckt diese Forschungsrichtung noch in den Kinderschuhen⁸⁸⁻⁹².

Parallel zur Cochrane Library wurde mit dem Aufbau eines Registers für eine systematische Erfassung und Bewertung von diagnostischen Tests, die Bayes Library of Diagnostic Studies and Reviews (www.bice.ch) begonnen. Viele methodische Probleme, wie z.B. die Frage des am besten geeigneten „Summary Measure“ von diagnostischen Tests sind noch unbeantwortet^{89; 93}. Auch bei der Betrachtung der einzelnen Tests lässt sich eine Weiterentwicklung feststellen. Neben Studien zu den Testqualitäten findet man vermehrt auch diagnostische „Outcome“-Studien, die den Informationsgewinn des Einzeltests im Kontext der Entscheidungskette untersuchen⁹⁴⁻⁹⁸.

1.2.4 Weitere Fragestellungen und Entwicklungen

Mit der stärkeren Einbindung der klinischen Epidemiologie in die Gestaltung und Steuerung des Gesundheitswesens werden vermehrt Fragen mit Systembezug an die klinische Epidemiologie gestellt. Bewertungen der Versorgungsqualität, Umsetzung und Sicherstellung einer hochwertigen Versorgung

in der Praxis, Messung der Auswirkungen außerhalb von Studiensituation. Der Bedarf der Entscheidungsträger an Kosten-Nutzen-Bewertungen oder Entscheidungsanalysen zu alternativen Vorgehensweisen steigt. Diese spannenden Entwicklungen auf der Steuerungsebene des Gesundheitssystems gehören jedoch nicht zum formulierten Schwerpunkt der vorliegenden Arbeit und sollen deshalb nicht berücksichtigt werden.

1.3 Von der klinischen Epidemiologie zur evidenzbasierten Medizin

1.3.1 Die Entwicklung zur evidenzbasierten Medizin

Mit ihrem Selbstverständnis als „basic science for clinical medicine“⁹⁹ war es der klinischen Epidemiologie von Anfang an ein zentrales Anliegen, ihr Wissen aus dem Elfenbeinturm der (klinischen) Forschung in die Versorgung hineinzutragen^{100; 101}. David Sackett an der McMaster University, Kanada und Robert und Susanne Fletcher in Chapel Hill, North Carolina integrierten die Methodik der klinischen Epidemiologie über die Aus- und Weiterbildung systematisch in die Versorgung¹⁰²⁻¹⁰⁴. In Yale wurde mit dem von Feinstein ins Leben gerufenen und über lange Jahre geleiteten *Robert Wood Johnson Clinical Scholarship Program* ein duales Weiterbildungsprogramm „Klinische Medizin / Klinische Epidemiologie“ mit hohem Renommee etabliert.

Die Benennung des neuen Konzepts als „Evidence-based Medicine“ zusammen mit der Ankündigung eines Paradigmenwechsels in der Medizin⁵⁵ führte in Nordamerika zu heftigen, zum Teil sehr emotional geführten Auseinandersetzungen. Mit der Gründung des „Centre for Evidence-Based Medicine“ in Oxford (1995) durch D. Sackett und M.Gray wurde die Debatte nicht weniger kontrovers in Europa fortgesetzt.¹⁰⁵⁻¹⁰⁹ In den Kritiken wurden häufig Probleme, wie der Mangel an robusten Daten, unzureichende Studienqualität, verzerrte Evidenzlage zugunsten medikamentöser Behandlungen oder die fragliche Übertragbarkeit der Studien auf den einzelnen Patienten angesprochen¹¹⁰. Dabei wurde und wird übersehen, dass ein Großteil der Missstände aus der EbM-Bewegung heraus aufgedeckt und benannt worden waren und viele Lösungskonzepte aus der EbM-Bewegung entwickelt wurden¹¹¹. Somit entstand und entsteht bisweilen der Eindruck, dass hier der Bote für die Nachricht geschlagen wird.

1.3.2 Was bedeutet „Evidenzbasierte Medizin“?

1.3.2.1 Die Definition

Darüber, was evidenzbasierte Medizin bedeutet, ist viel gestritten worden. Im Kontext der vorliegenden Arbeit beziehe ich mich auf die Sichtweise des Arztes in der Patientenversorgung, wie sie von D.L. Sackett auf dem Höhepunkt der Auseinandersetzungen in Großbritannien formuliert worden war¹¹²:

„Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external evidence from systematic research.“

In dieser Definition ist eine transparente und explizite Vorgehensweise in der Patientenversorgung ein wesentliches Merkmal der EbM. Das Regelwerk dafür, was „Evidenz“ (abgeleitet vom angelsächsischen „evidence“, im Sinne von Nachweis, Beweis) bedeutet, stammt aus der patientenorientierten Forschung, deren Aussagen nach den Kriterien der klinischen Epidemiologie auf ihre Glaubwürdigkeit hin überprüft werden. Für eine „evidenzbasierte“ Entscheidungsfindung für den einzelnen Patienten benötigt der Arzt weitere Fertigkeiten, um dessen individuelles Krankheitsprofil zuverlässig einzuschätzen und seine Bewertungen von Risiko und Prognose mit den Ergebnissen der klinischen Forschung zu verbinden. Das erfordert ein begründetes Abwägen von Nutzen und Schaden, von Kosten und persönlichen Werten und Präferenzen. Um die Vorstellungen und Werte der Patienten in die Entscheidungsfindung zu integrieren, werden Methoden benötigt, mit denen sich Informationen über Behandlungsnutzen und Schadensrisiko, Ausgangsrisiken und Prognose dem Patienten verständlich kommunizieren lassen, eine Forschungsrichtung, die derzeit noch in ihren Kinderschuhen steckt¹¹³.

In der Medizin existieren sehr viele Grauzonen¹⁰⁷. Eine Praxis nach den Kriterien der EbM bedeutet, diese Graubereiche zu kennen, mit der daraus resultierenden Unsicherheit bewusst umzugehen und die Wissenslücken mit klinischer Urteilsfähigkeit und professioneller Erfahrung zu kompensieren.

Somit wird klar, dass EbM keine uniforme, für alle Situationen in gleicher Weise geltende Lösung anbietet, aber eine Voraussetzung für eine rationale Entscheidungsfindung ist. Selbst rationale

Entscheidungen werden jedoch von anderen Aspekten wie Interessen und Überzeugungen, Prioritäten und Zwängen, vorhandenen oder fehlenden Ressourcen erheblich mitbeeinflusst. Die differenzierte Bestandsaufnahme der Evidenzlage stellt somit ein wichtiges Werkzeug für eine rationale Entscheidungsfindung dar, aber die Evidenzlage allein kann keine Entscheidungen fällen und die evidenzbasierte Medizin erhebt auch nicht diesen Anspruch. Um den immer öfter anzutreffenden Missbrauch der EbM als scheinbar „oberste Instanz“ entgegenzutreten, sollte auch bevorzugt von „einer“ evidenzbasierten Medizin statt von „der“ Evidenzbasierten Medizin gesprochen werden.

1.3.2.2 Der Evidenzpfad - Die Bedeutung der evidenzbasierten Medizin für das Gesundheitswesen

Mit dem Evidenzpfad (Abb. 1.1) lässt sich das Konzept eines evidenzbasierten Gesundheitswesens mit seinen vielfältigen Bestandteilen beschreiben.

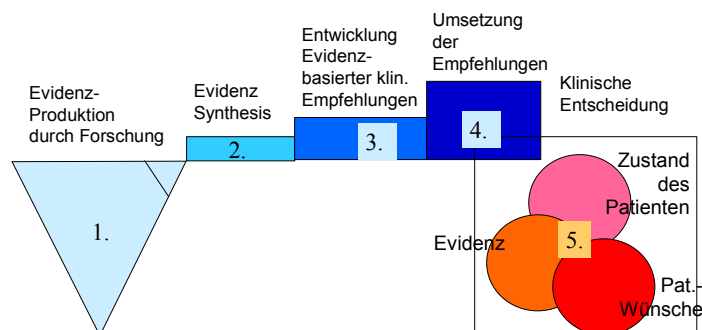


Abb. 1.1: Der Evidenzpfad – Von der Evidenzproduktion zur Anwendung

Dabei überbrückt der Pfad mit den einzelnen Schritten - von der Evidenzproduktion (Primärforschung), über die Evidenzsynthese (Cochrane Collaboration, Bayes Collaboration, HTA-Berichterstattung), die Entwicklung von evidenzbasierten Policies (insbesondere Leitlinien und Disease-Management-Programmen) und deren Umsetzung bis hin zur Patientenbetreuung - die vielfach beschriebene Kluft zwischen der klinischen Forschung und der Anwendung in der

Patientenversorgung. Die drei ineinandergreifenden Säulen der Patientenversorgung in dem abgeschlossenen Quadrat demonstrieren, wie bei einer evidenzbasierten Entscheidungsfindung neben der Evidenzlage aus Studien die Einschätzung und Bewertung der Patientensituation durch den Arzt sowie die Wünsche und Präferenzen des Patienten berücksichtigt werden müssen, um der Komplexität und den Besonderheiten eines jeden Patienten gerecht zu werden.

1.3.3 EbM in Deutschland

In Deutschland wurde die evidenzbasierte Medizin und ihr Potential für die Patientenversorgung wie für das Gesundheitswesen erstmalig 1996 sehr ausführlich und differenziert vorgestellt¹⁰⁰. Obwohl seither weniger als sieben Jahre vergangen sind, ist das Konzept der EbM, wenn auch oft nur als Schlagwort, in nahezu alle Bereiche des Gesundheitswesens eingedrungen. Insbesondere auf der Systemebene werden die Ansprüche von EbM zunehmend als Messlatte (Gesetzgebung, Bundesausschuss der Ärzte und Krankenkassen) eingesetzt. Über die Vorgabe zur Erstellung von evidenzbasierten Leitlinien wurden die Meinungsbildner in den medizinischen wissenschaftlichen Fachgesellschaften unmittelbar mit dem Thema konfrontiert. Mit der Gründung des Deutschen Netzwerks EbM wurde eine breite und offene Plattform für Zusammenarbeit und Austausch geschaffen. So entstand in Zusammenarbeit von Ärztlicher Zentralstelle Qualitätssicherung und dem Deutschen Netzwerk EbM das mehrstufige Curriculum EbM (Kapitel 4), mit dem sich die regionalen EbM-Kurse quer durch Deutschland eine gemeinsame Grundlage erarbeitet haben. Es gibt Projekte zur Integration von EbM in die Lehre (Herdecke, Ulm, München, Freiburg) und Modellversuche, EbM in den Klinikalltag zu integrieren (Wuppertal¹¹⁴, Berlin¹¹⁵, Köln, St. Franziskus Hospital).

Bei den Entscheidungsträgern hat sich bezüglich EbM eine gewisse Ernüchterung breitgemacht: Die Komplexität der EbM ist von vielen stark politisch motivierten Befürwortern erheblich unterschätzt worden. Den (falschen) Erwartungen, eine einfache autoritative Antwort auf komplexe gesundheitspolitische Fragen zu liefern, kann und will die EbM nicht entsprechen. Zusätzlich existiert in Deutschland bezüglich des notwendigen methodischen Unterbaus ein eklatanter Mangel. Es fehlt eine Infrastruktur, wie z.B. ein EbM-Zentrum oder eine ausreichende und ausreichend langfristige finanzielle Ausstattung vorhandener Strukturen, wie dem Deutschen Cochrane Zentrum, die die

notwendigen Dienstleistungen erbringen könnten. Und nicht zuletzt fehlen Anreizsysteme für die Ärzteschaft, sich mit der evidenzbasierten Medizin auseinanderzusetzen. Anforderungen und Realität klaffen auseinander.

Einige dieser Aspekte werden in den folgende Kapiteln systematisch untersucht und Lösungsansätze entwickelt.

2 Beobachtung oder Experiment? Empirische Methodenforschung

Aufbau des Kapitels

In der klinischen Epidemiologie existiert zwischen den „Trialists“ und den „Observationalists“ ein grundlegender Dissens über die Frage, ob die Randomisierung im klinischen Experiment eine notwendige Voraussetzung für eine zuverlässige Effektschätzung darstellt. Diese Frage wurde von unserer Arbeitsgruppe (R. Kunz, Berlin und A. Oxman, G. Vist, Oslo) erstmalig anhand realer Daten aus der klinischen Forschung systematisch untersucht (**The Unpredictability Paradox; 2.1**). Die dadurch ausgelöste Diskussion wurde in den nachfolgenden Jahren in hochrangigen Journals mit weiteren empirischen Studien (Benson *N Engl J Med* 2000; Concato *N Engl J Med* 2000; Ioannidis *JAMA* 2001, Kunz, *Cochrane Library* 2002) fortgesetzt. Diese Diskussion wird zusammenfassend wiedergegeben und kritisch bewertet (**2.2**). Das Kapitel schließt mit der Beschreibung von der Entwicklung und dem Aufbau eines Registers für empirische methodische Studien durch unsere Arbeitsgruppe sowie unseren Untersuchungen zur Entwicklung systematischer effizienter Suchstrategien für relevante Studien in gängigen Datenbanken (**2.3**). Das Register wurde später in die *Cochrane Library* (*Cochrane Review Methodology* Datenbank) integriert und stellt eine fundamentale Grundlage für die empirische Methodikforschung dar.

Eigene eingeschlossene Publikationen zum Thema

- Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *The Cochrane Library* (3). 2002.
- Kunz R, Vist G, Neumayer HH, Oxman AD. Further evidence on the need for randomised controlled trials (Abstract). 9th Annual Cochrane Colloquium, Lyon 2001.
- Kunz R, Khan KS, Neumayer HH. Observational studies and randomized trials. *N.Engl.J Med* 2000; 343: 1194 - 5.
- Kunz R, Oxman AD. Two systematic reviews - two different answers? *BMJ* 1999. bmj.com/cgi/eletters/319/7205/312#4411
- Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.
- Kleijnen J, Gotzsche PC, Kunz R, Oxman AD, Chalmers I. So, what's so special about randomisation? In Maynard A, Chalmers I, eds. *Non-random Reflections on Health Services Research*, pp 93-106. London: BMJ Publishing Group, 1997.

2.1 Beobachtung oder Experiment – ein empirischer Vergleich (eigene Untersuchungen)

Evidenzstufen („Levels of Evidence“)¹¹⁶ und Empfehlungsstärken („Grades of Recommendations“^{116,117}) für Bewertungen von Therapien und Interventionen gehören zu den Kernkonzepten der evidenzbasierten Medizin. Die Evidenzstufen beschreiben eine Hierarchie der Studiendesigns auf der Grundlage ihrer Fähigkeit, den wahren Effekt einer Intervention zuverlässig und mit möglichst geringem Anteil an Irrtum (Bias oder Confounding) zu schätzen. Schutz vor Irrtümern bieten sorgfältig ausgewählte und konsequent eingesetzte methodische Verfahren in der Planung, Durchführung und Analyse von Studien. Designs mit mehr Vorkehrungen vor potentiellen Irrtumsquellen werden in dieser Hierarchie höher angesiedelt als Designs mit größerem Irrtumspotential und höherem Risiko für fehlerhafte Ergebnisse. Die Hierarchie der Studiendesigns¹¹⁸ entstand auf der Grundlage von klinisch-epidemiologischen Regeln („Rules of Evidence“), also auf logischen Überlegungen, unterstützt von anekdotischen Einzelberichten, wo fehlende protektive Strategien zu fatalen Fehleinschätzungen oder Irrtümern geführt haben^{39; 119; 120}. Die ersten Listen für einen Qualitätscheck umfassten mehr als 30 methodische Merkmale in der Planungs-, Durchführungs- oder Analysephase¹²¹, andere folgten⁷⁵. Bei den Interventionsstudien erhalten randomisierte klinische Experimente (Evidenzstufe I) eindeutig eine höhere Wertschätzung als nichtrandomisierte Beobachtungsstudien (Evidenzstufe II und niedriger). Doch wie groß ist der Schutz, den die einzelnen methodischen Verfahren liefern? Welche Evidenz rechtfertigt diese Einstufung? Und hat die Beantwortung dieser Frage praktische Konsequenzen?

Durch die zunehmende Verfügbarkeit riesiger Datensätze mit Patienteninformationen bei den Krankenkassen, in elektronischen Patientenakten oder Forschungsdatenbanken gewinnt die Frage „Beobachtung oder Experiment“ zur Erfassung der Wirksamkeit medizinischer Maßnahmen erneut an Bedeutung. Trotz stichhaltiger Gründe für eine stärkere Erschließung dieser Datenquellen (einfacher, kostengünstiger Zugang zu umfassenden Patientenkollektiven, Abbildung breiter Bevölkerungsgruppen, Möglichkeiten für ergiebige Untergruppenanalysen, lange Nachbeobachtungszeiten zur Erfassung seltener Ereignisse) bleibt die Frage: Liefern die Vergleiche aus (meist) retrospektiven Beobachtungen valide Ergebnisse, d.h. werden Effekte, die wirklich

existieren, in der wahren Effektgröße auch abgebildet bzw. Effekte, die nicht vorhanden sind, auch nicht fälschlicherweise dargestellt. Ein wichtiges Argument gegen einen unkritischen Nutzen von Datenbankanalysen ist die Gefahr der Selektionsbias. „In medicine, the doctor chooses the therapy precisely in order to affect outcomes“¹²², weshalb diese Populationen besonders dazu neigen, sich zu unterscheiden. Auch wenn die Protagonisten von Beobachtungsstudien auf die Weiterentwicklung in Methodik und Analysetechnik hinweisen¹²³, können Defizite in der Dokumentation vorhandener Unterschiede und in dem begrenzten medizinischen Wissen über ausschlaggebende Determinanten und prognostische Faktoren der meisten Erkrankungen damit nicht korrigiert werden. Lange haben „Observationalists“ wie „Trialists“ ihre Positionen mit passenden Beispielen unterstrichen¹²⁴⁻¹²⁸, aber die Diskussion damit nicht weitergebracht.

Wir haben uns der Frage, Beobachtung oder Experiment, mit der Methodik der systematischen Übersichtsarbeit genähert, in dem wir das methodische Verfahren „Randomisierung“ bzw. „keine Randomisierung“ als Intervention aufgefasst und die resultierenden Effektgrößen miteinander verglichen haben. Um aus dem Dilemma der anekdotischen Einzelevidenz herauszukommen, mit der sich nahezu jede Hypothese belegen lässt, haben wir als „Population“ Studien gewählt, die unsere Fragen an einem größeren, möglichst systematisch zusammengestelltem Kollektiv von Studien untersucht haben und diese Ergebnisse auf der Meta-Ebene verglichen. Unsere Studie (1998 im BMJ publiziert) war die erste Studie mit diesem Ansatz und eine der ersten Arbeiten in der empirischen Methodikforschung. Thema und Methodik wurden von anderen Arbeitsgruppen aufgegriffen und die Diskussion in hochrangigen Journals (N Engl J Med, JAMA, s.unten) fortgesetzt.

Im Folgenden soll unsere Studie zusammengefasst dargestellt werden:

2.1.1 Das „Unpredictability Paradox“ – Randomisierte und nichtrandomisierte Studien im Vergleich

2.1.1.1 Fragestellungen

In unserer Studie wurden vier Vergleiche untersucht: Wie verhält sich die Effektgröße

- 1) bei randomisierten verglichen mit nichtrandomisierten Studien mit gleichen Interventionen
- 2) bei randomisierten verglichen mit nichtrandomisierten Studien mit heterogenen Interventionen
- 3) bei angemessen maskierter verglichen mit nicht angemessen maskierter Randomzuteilung
- 4) bei qualitativ hochwertigen verglichen mit weniger hochwertigen RCTs, bei denen die Randomisierung als methodisches Verfahren nicht von den Auswirkungen anderer methodischer Verfahren getrennt werden konnte. (Hochwertig bedeutet hier den Einsatz einer größeren Anzahl methodischer Manöver, die aufgrund von experimentellen Untersuchungen (z.B. Jadad-Score) oder logischen Argumenten für die Fragestellung als notwendig erachtet und a priori festgelegt worden waren).

2.1.1.2 Methodik

2.1.1.2.1 Literatursuche und Auswahl der Studien

Wir durchsuchten folgende Quellen nach geeigneten Studien: Cochrane Review Methodology Datenbank, methodologische Bibliographien (z.B. Controlled Clinical Trials Bibliographie), Medline, Science Citation Index (über Indexartikel), Handsearching methodischer Journals (Controlled Clinical Trials, Journal of Clinical Epidemiology, Statistics in Medicine), persönliche Kommunikation mit Methodikern, Referenzlisten. Eine grosse Anzahl an Studien war als Nebenprodukt der Suchstrategien zur Entwicklung der Datenbank für methodische Studien (Kapitel 2.3) aufgefunden worden. Klassische Beobachtungsstudien (Fall-Kontroll-Studien, Kohortenstudien) und Einzelbeispiele („anecdotal evidence“) waren ausgeschlossen. Die Suche wurde bis Sommer 1998 durchgeführt. Alle potentiell relevanten Studien wurden in der Druckversion von zwei Autoren (R.K. und A.O.) unabhängig gesichtet und ausgewählt. Bei fehlender Übereinstimmung wurde Ein- bzw. Ausschluss ausdiskutiert.

2.1.1.2.2 Methodische Bewertung

Die methodische Bewertung der Kohorten / Meta-Analysen erfasste sechs Items (Auswahl der Studien anhand expliziter Kriterien durch mehrere Wissenschaftler mit Dokumentation der Übereinstimmung („Agreement“), fortlaufende bzw. komplette Studienstichprobe, Kontrolle für andere methodische Verfahren, z.B. doppelte Verblindung oder vollständige Nachbeobachtung und für Heterogenität in den Studienpopulationen / Interventionen der einzelnen Studien, Untersuchung vergleichbarer Endpunkte). Die Studienqualität wurde einer der drei Kategorien zugeordnet („keine bedeutsamen Mängel“, „potenziell bedeutsame Mängel vorhanden“, „folgenreiche Mängel vorhanden“) und ging in die Gesamtbewertung ein.

2.1.1.2.3 Datenextraktion

Für jede Übersicht wurden die relevanten Daten (Stichprobe der ausgewählten Studien, durchgeführte Vergleiche, Ergebnisse, methodische Bewertung) von der Autorin extrahiert und von einem zweiten Wissenschaftler (Andy Oxman (A.O.)) gegengelesen. Die beobachtete Beziehung zwischen Randomisierung und geschätztem Behandlungseffekt (Über- oder Unterschätzung, kein Unterschied, Umkehreffekt) wurde dokumentiert und die Heterogenität innerhalb und zwischen den Studien untersucht.

2.1.1.2.4 Analyse

Die Analyse wurde im Wesentlichen deskriptiv durchgeführt, wobei wir die Schätzwerte der randomisierten Studien (bzw. von adäquat randomisierten Studien, von Studien mit mehr protektiven methodischen Verfahren, ...) als Referenzstandard für die Schätzwerte der nichtrandomisierten Studien definierten.

2.1.1.3 Ergebnisse

Insgesamt wurden 18 Kohorten aus Studien und Meta-Analysen mit insgesamt 1211 einzelnen klinischen Studien untersucht^{17; 127-142}.

2.1.1.3.1 Methodische Bewertung

Die eingeschlossenen Arbeiten zeigten ausgeprägte Variabilität in ihrer Qualität (Tab. 2.1): Nur vier Arbeiten erfüllten alle vorgegebenen Qualitätskriterien^{137; 139-141}, drei davon waren Untergruppenanalysen, die mit dem Vergleich „Hochwertige verglichen mit weniger hochwertigen Studien“ die Robustheit der Ergebnisse ihrer Meta-Analyse testeten. Die restlichen Studien wiesen wegen unzureichender Berücksichtigung anderer methodischer Verfahren^{17; 129; 135; 136} oder klinischer Heterogenität besonders bei Populationen und Interventionen potentiell bzw. sicher schwerwiegende Mängel auf^{17; 127; 129; 134-136; 142; 143}.

2.1.1.3.2 RCTs verglichen mit Nicht-RCTs bei vergleichbaren Interventionen

Tabelle 2.2 beschreibt die acht Studien^{127; 128; 130-133; 144}, in denen die Effektgrößen von RCTs und Nicht-RCTs miteinander verglichen wurden. In fünf von acht Studien fanden die Nicht-RCTs größere Effekte. Dabei kamen die Behandlungsgruppen häufig zu vergleichbaren Ergebnissen, aber der Behandlungseffekt wurde durch die deutlich schlechteren Ergebnisse der nichtrandomisierten (oft historischen) Kontrollen fälschlicherweise aufgebläht. Eine Studie zeigte für beide Zuordnungsmethoden vergleichbare Ergebnisse und zwei Studien fanden in den Nicht-RCTs kleinere Behandlungseffekte, wobei die Patienten der nichtrandomisierten Behandlungsgruppe gleichzeitig eine schlechtere Prognose hatten (eine Studie). Im Vergleich schwankte der Behandlungseffekt von Nicht-RCTs zu RCTs von einer Unterschätzung des Effekts von 76% bis hin zu einer Überschätzung um 160%.

2.1.1.3.3 RCTs verglichen mit Nicht-RCTs bei heterogenen Interventionen

Die Interpretation der drei Studien mit heterogenen Interventionen und unterschiedlichen Studiendesigns (RCTs und Nicht-RCTs, Cross-over-Designs und Beobachtungsstudien)^{135; 136; 138} gestaltet sich wesentlich schwieriger (Tab. 2.3): Um für die unterschiedlichen klinischen Endpunkte der Einzelstudien eine gemeinsame Größe zu erhalten, konvertierten die Autoren in allen drei Arbeiten die unterschiedlichen Endpunkte in ein standardisiertes Effektmaß. Zwischen Studiendesign / -qualität und dem standardisierten Effektmaß ließ sich keine konsistente Beziehung darstellen. Möglicherweise entstand durch die ausgeprägte klinische Heterogenität der einzelnen Studien und zahlreichen anderen

Faktoren, die nicht berücksichtigt wurden, so viel „Rauschen“, dass potentiell vorhandene Assoziationen zwischen Randomisierung und Effektgröße nicht oder nur verzerrt abgebildet wurden.

2.1.1.3.4 Angemessen maskierte verglichen mit nicht angemessen maskierter Randomisierung

Ziel der verdeckten Randomisierung zu einer Behandlung ist, gegen Verzerrungen bei der Gruppenzuteilung zu schützen. Chalmers zeigte in zwei sehr frühen Studien, dass selbst innerhalb von randomisierten Studien eine nicht adäquat verblindete („maskierte“) Gruppenzuteilung die prognostischen Faktoren der Studien so fehlverteilt, dass die Behandlungseffekte bis zum siebenfachen überschätzt wurden^{17; 127}. Hauptproblem dieser frühen Arbeit sind die fehlenden Kontrollen für andere methodische Verfahren. Doch 15 Jahre später wurden die Ergebnisse bestätigt, als Schulz nach Korrektur für Verblindung und Nachbeobachtung mittels multi-variater Analyse ähnliche, wenn auch weniger ausgeprägte Effekte nachweisen konnte¹³⁷: Verglichen mit maskierter Zuordnung fanden sich bei inadäquater Gruppenzuordnung im Durchschnitt um 40% aufgeblähte Effekte (Odds Ratio).

2.1.1.3.5 Hochwertige verglichen mit weniger hochwertigen RCTs

Die vierte Kategorie waren Sensitivitätsanalysen innerhalb von Meta-Analysen über RCTs zu spezifischen Behandlungsformen, mit denen die Stabilität des gefundenen Effekts überprüft wurde. Ein Vergleich von hochwertigen RCTs mit weniger hochwertigen RCTs (und größerem Potential für Bias) zeigte bei den Effektgrößen erhebliche Unterschiede^{129; 139; 140; 142; 145} in beide Richtungen (Über- wie Unterschätzung des Effekts). In einer Meta-Analyse über eine spezifische Infertilitätsbehandlung fand sich sogar ein qualitativer Unterschied, das heißt, dass in hochwertigen Studien ein schädigender Effekt nachgewiesen wurde (Rückgang der Schwangerschaften: OR 0.5), während sich in den methodisch minderwertigen Studien ein positiver Effekt (Anstieg der Schwangerschaften: OR 2.6)¹⁴⁵ zeigte. In zwei Meta-Analysen unterschätzten die minderwertigen RCTs den günstigen Effekt der Intervention um 27% bis 100%. Konsequenterweise wäre damit durch die Ergebnisse minderwertiger RCTs eine eigentlich wirksame Intervention verworfen worden.

Tab. 2.1: Methodische Qualität der einzelnen Übersichtsarbeiten / Kohorten von Studien

Study	Sample	Confounding control?	Reproducibility	Outcome measure	Overall judgement
	Consecutive or complete sample?	1. Other methodological criteria: 2. Participants/clinical interventions	1. Explicit description 2. Assessment of agreement		
Imperiale 1990	Yes	1 Yes 2 Yes	1 Yes 2 Yes	Same outcome measure	No important flaw
Khan 1996	Yes	1 Yes 2 Yes	1 Yes 2 partly	Same outcome measure	No important flaw
Nurmohamed 1992	Yes	1 Yes 2 Yes	1 Yes 2 partly	Same outcome measure	No important flaw
Schulz 1995	Yes	1 Yes 2 Yes	1 Yes 2 partly	Same outcome measure	No important flaw
Ortiz 1998	Yes	1 Yes 2 No	1 Yes 2 Yes	Same outcome measure	No - possibly important flaw
Carroll 1996	Yes	1 Partly 2 No	1 partly 2 No	p-value or vote counting	Possibly important flaw
Chalmers 1983	Yes	1 No 2 Partly	1 Yes 2 No	Same outcome measure	Possibly important flaw => overestimation
Diehl 1986	No	1 No 2 Yes	1 Yes 2 No	Same outcome measure	Possibly important flaw -- direction of bias unclear
Emerson 1990	No	1 Yes 2 No	1 Yes 2 partly	Same outcome measure	Possibly important flaw => underestimation
Reimold 1992	Yes	1 Partly 2 Partly	1 Yes 2 No	Same outcome measure	Possibly important flaw
RMIT Group 1994	Yes	1 partly 2 Yes	1 Yes 2 partly	Same outcome measure	Possibly important flaw
Watson 1994	Yes	1 Partly 2 Partly	1 Yes 2 No	Same outcome measure	Possibly important flaw
Colditz 1989	Yes	1 Partly 2 No	1 Yes 2 partly	Standardized outcome measure	Possibly important - major flaw =>underestimation
Miller 1989	Yes	1 Partly 2 No	1 Yes 2 partly	Standardized outcome measure	Possibly important - major flaw =>underestimation
Ottensmeyer 1992	No	1 No 2 No	1 Yes 2 Yes	Standardized outcome measure	Possibly important - major flaw =>any direction
Pyorala 1995	Yes	1 No 2 No	1 partly 2 No	Same outcome measure	Possibly important - major flaw
Chalmers 1977	Yes	1 No 2 No	1 Yes 2 No	Same outcome measure	Major flaw => overestimation
Sacks 1982	No	1 No 2 Partly	1 Yes 2 No	p-value or vote counting	Major flaw => overestimation

Tab. 2.2: RCTs verglichen mit Nicht-RCTs bei Übersichten zu gleichen Interventionen

Study	Sample (search strategy)	Comparison	Results	Direction of bias
Chalmers 1977	32 controlled studies of anticoagulation in acute myocardial infarction (systematic)	RCTs with CCTs and HCTs on case fatality rate, rate of thromboembolism, and haemorrhages	Relative risk reduction for mortality overestimated by 35% in HCTs and 6% in CCTs compared with RCTs. Case fatality rate highest in HCTs (38.3%) compared with RCTs (19.6%) and CCTs (29.2%). Similar pattern for thromboembolism	Overestimation of effect
Sacks 1982	Sample of 50 RCTs and 56 HCTs, assessing 6 interventions (treatment of oesophageal varices, coronary artery surgery, anticoagulation in myocardial infarction, chemotherapy for colon cancer and melanoma, and diethylstilboestrol for recurrent miscarriage) (at hand)	RCTs with HCTs on frequency of detecting statistically significant results ($p=0.05$) of primary outcome and reduction of mortality	20% of the RCTs found a statistically significant benefit from the new treatment compared with 79% of the HCTs. Relative risk reduction of mortality in HCTs v RCTs was 0.49/0.27 (1.8) for cirrhosis, 0.68/0.26 (2.6) for coronary artery surgery at 3 years, 0.49/0.22 (2.2) for anticoagulation in myocardial infarction, and 0.67/0.02 for diethylstilboestrol in recurrent miscarriage. Outcomes in treatment groups were similar in both designs, but outcomes in control groups were worse among historical controls	Overestimation of effect
Diehl 1986	19 RCTs and 17 HCTs for 6 types of cancer (breast, colon, stomach, lung cancer, melanoma, soft tissue sarcoma) (reference lists of two textbooks)	Matching of randomised and historical controls for disease, stage, and follow up, and comparison on survival and relapse free survival 18 of 43 matched control groups (42%) varied by >10% (absolute difference in either outcome), 9 (21%) by >20%, and 2 (5%) by >30%.	Survival or relapse free survival was better in RCTs compared with HCTs in 17/18 matches	Overestimation of effect
Reimold 1992	6 RCTs and 6 CCTs of chinidine in atrial fibrillation (systematic)	RCTs and CCTs on maintenance of sinus rhythm 3, 6, and 12 months after cardioversion	At 3 months, beneficial effect of maintaining sinus rhythm with chinidine was 54% less in non-RCTs compared with RCTs, and was 76% less at 12 months	Underestimation of effect
Recurrent Miscarriage Immunotherapy Trialists Group 1994	9 RCTs and 6 CCTs (with self selected treatment) of allogenic leucocyte immunotherapy for recurrent miscarriage (systematic)	RCTs and CCTs on live birth rate	Beneficial effect of immunotherapy on birth rate among pregnant women was 9% larger in CCTs compared with RCTs, but was 63% lower in CCTs when all women were considered	Underestimation of effect when all women considered, similar effect for pregnant women
Watson 1994	4 RCTs and 6 CCTs/HCTs of oil soluble contrast media during hysterosalpingography in infertile couples (systematic)	RCTs and CCTs/HCTs on pregnancy rate	RCTs and CCTs/HCTs detected similar increases in pregnancy rates: odds ratio for RCTs 1.92 (95% CI, 1.33 to 2.68) and for CCTs/HCTs 1.92 (1.55 to 2.38)	Similar effect
Pyörälä 1995	11 RCTs and 22 (not further specified) non-RCTs on hormonal therapy in cryptorchidism (systematic)	RCTs and non-RCTs on the descent of testes after therapy with luteinising hormone releasing hormone or human chorionic gonadotrophin	Success rate of descent of testes after therapy with luteinising hormone releasing hormone was 2.3 times larger in non-RCTs than in RCTs and 1.7 times larger after therapy with human chorionic gonadotrophin	Overestimation of effect
Carroll 1996	17 RCTs and 19 non-RCTs (including HCTs or trials with inadequate randomisation procedures) on transcutaneous electrical nerve stimulation (systematic)	RCTs and non-RCTs on control of postoperative pain	Transcutaneous electrical nerve stimulation judged ineffective at improving postoperative pain in 85% of RCTs, while 89% of non-RCTs concluded that it did improve postoperative pain	Overestimation of effect

RCT=Randomised controlled study; non-RCT=non-randomised controlled study; CCT=concurrently controlled trial; HCT=historically controlled trial.

Tab. 2.3: RCTs verglichen mit Nicht-RCTs zu Übersichten mit Vergleichen von heterogenen Interventionen

Study	Sample (search strategy)	Comparison	Results	Direction of bias
Colditz 1989	113 studies published in 1980 comparing new interventions with old, identified in leading cardiology, neurology, psychiatry, and respiratory journals (systematic)	36 parallel RCTs, 29 randomised COTs, 46 non-randomised COTs, 3 CCTs, 5 ECTs, 9 observational studies compared for "treatment gain" (Mann-Whitney statistic), and relation between quality score and "treatment gain" assessed	All but one design achieved similar "treatment gains" (0.56-0.65). Overall, 89% of new treatments were rated as improvements, but only non-randomised COTs detected a significantly higher "treatment gain" from the new treatment compared with RCTs (p=0.004). Within RCTs, there was no correlation between quality score and "treatment gain" (p=0.18)	Inconclusive
Miller 1989	188 studies comparing new surgical interventions with old, published in 1983 and identified in leading surgical journals (systematic)	81 RCTs, 15 CCTs, 27 HCTs, 91 observational studies, 7 BASs compared on "treatment gain" (Mann-Whitney), and association between treatment success and study design and the relation between quality score and treatment gains assessed	Non-significant trend towards larger "treatment gains" for new treatments on the principal disease in non-RCTs (0.56 to 0.78) than in RCTs (0.56). For treatment of complications the "treatment gain" was similar across all study designs (0.54 to 0.55) except in BASs (0.90). Within RCTs, there was no correlation between quality scores and treatment gains (p=0.7)	Inconclusive
Ottensmeyer 1992	Sample of 30 RCTs and 30 trials with non-random process of allocation, eg matching or HCTs (systematic search of N Engl J Med and JAMA across several medical specialties)	RCTs and non-RCTs on treatment effects as measured by standardised mean differences	No difference in treatment effect found between non-RCTs (0.23) and RCTs (0.21)	Similar effects

RCT=Randomised controlled study; non-RCT=non-randomised controlled study; CCT=concurrently controlled trial; HCT=historically controlled trial.COT=Crossover trial; CCT=concurrently controlled trial; ECT=external control study; BAS=before and after study; HCT=historically controlled trial.

Tab. 2.4: Experimentelle Studien mit angemessen maskierter Zuordnung verglichen mit unangemessen maskierter Zuordnung

Study	Sample (search strategy)	Comparison	Results	Direction of bias
Chalmers 1983	145 controlled trials of treatment for acute myocardial infarction (systematic)	Studies with different allocation schemes (non-random, non-concealed random, and concealed random allocation) on maldistribution of prognostic variables, frequency of significant outcomes, and case fatality rates	In non-RCTs, non-concealed RCTs, and RCTs with concealed allocation, the maldistribution of prognostic factors was 34%, 7%, and 3.5% respectively, frequency of significant outcomes was 25%, 11%, and 5% respectively, average relative risk reduction for mortality was 33%, 23%, and 3% respectively. Case fatality rate for control groups was 32%, 23%, and 16% and for treatment groups was 21%, 18%, and 16% respectively	Overestimation of effect
Schulz 1995	250 RCTs from 33 meta-analyses (Cochrane Pregnancy and Childbirth Database)	Association between methodological features of controlled trials (allocation concealment, double blinding, and follow up), and treatment effect (odds ratio)	Treatment effect overestimated by 41% in RCTs with inadequate concealment and by 30% in RCTs with unclear adequacy of concealment compared with those with adequate concealment (p=0.001) after adjustment for other methodological features. Studies with no double blinding overestimated treatment effect by 17% compared with double blinded studies (p=0.01). Lack of complete follow up had no influence on treatment effect (7%, p=0.32)	Overestimation of effect

RCT=Randomised controlled trial

Tab. 2.5: Qualitativ hochwertige verglichen mit weniger hochwertigen Studien

Study	Sample (search strategy)	Comparison	Results	Direction of bias
Emerson 1990	Sample of 7 meta-analyses with 107 primary studies where full information about quality scores was available (at hand)	Assessment of relation between quality score and (a) observed treatment difference and (b) variation of observed treatment difference	No correlation detected between either quality score and treatment difference or variation of treatment difference within each meta-analysis or in combined analysis (p=0.29)	Similar effects
Imperiale 1990	Meta-analysis of 11 RCTs of steroids in alcoholic hepatitis (systematic)	Short term mortality in studies with high and low methodological quality	In studies with low quality, relative risk reduction on mortality was 86% smaller than the reduction observed in high quality studies. In studies with low quality and hepatic encephalopathy no effect was observed, while the relative risk reduction of mortality in high quality studies was 55%	Underestimation of effect
Nurmohamed 1992	Meta-analysis of 35 surgical and orthopaedic RCTs on low molecular weight heparin as thromboprophylaxis (systematic)	Relative risk reduction for deep vein thrombosis and pulmonary embolism in studies of high and low methodological quality	In studies with low quality, relative risk reduction for venous thrombosis in surgical trials was 2.6 times larger, and in orthopaedic trials 1.4 times larger, than studies with high quality. Relative risk reduction for pulmonary embolism in surgical trials was 1.7 times larger, and in orthopaedic trials 2.8 times larger, than studies with high quality	Overestimation of effect
Khan 1996	Meta-analysis of 9 RCTs (parallel or crossover design) evaluating the effect of anti-oestrogen treatment in male infertility (systematic)	Pregnancy rates in studies with high and low methodological quality	In studies of low quality, pregnancy rate increased under treatment (odds ratio 2.6), but declined under treatment in high quality studies (0.5)	Reversal of effect
Ortiz 1998	Meta-analysis of 7 RCTs on the effect of folic or folinic acid v placebo (systematic)	Frequency of gastrointestinal side effects in studies with high and low methodological quality	In studies with low quality there was a 43% reduction in the odds ratio of side effects (0.57) compared with a 70% reduction in studies with high quality (0.3)	Underestimation of effect

RCT=Randomised controlled trial.

2.1.1.4 Diskussion

Die in den Review eingeschlossenen Übersichten zeigen, dass nichtrandomisierte bzw. nicht angemessen randomisierte Studien (im Vergleich zu (angemessen) randomisierten) Studien dazu neigen, den Effekt einer Behandlung zu überschätzen. Dieses Phänomen trat jedoch nicht bei allen Vergleichen auf, vielmehr fanden sich bei einem nicht unerheblichen Prozentsatz in den nichtrandomisierten Studien gleich große oder kleinere Effekte. In einigen Beispielen wurde sogar ein Effekturnkehr beobachtet.

Der „wahre“ Effekt von Interventionen ist immer unbekannt und auch hochwertige Studien können nur eine Schätzung des wahren Effekts liefern. Im Methodikteil unserer Studie haben wir die randomisierten Studien als Referenzwert für die Bewertung der nichtrandomisierten Studien festgelegt. Diese Entscheidung impliziert nicht, dass der „wahre“ Effekt bekannt ist oder dass die Schätzungen von RCTs immer näher an dem unbekanntem „wahren“ Wert liegen als die von Nicht-RCTs. Auch in randomisierten Studien, unserem „Goldstandard“, können Mängel auftreten. Allerdings liegt unseren Überlegungen die Prämisse der gesamten (epidemiologischen) Forschung zugrunde, dass Studien mit größerem Schutz vor systematischen Fehlern mit höherer Wahrscheinlichkeit nahe dem unbekanntem „wahren“ Wert zu liegen kommen als Studien mit weniger protektivem Design.

Dennoch weist unsere Untersuchung einige Einschränkungen auf: Trotz umfangreicher Bemühungen (s. 2.3) ist es nicht gelungen, eine umfassende Suchstrategie aufzubauen, die in den gängigen Datenbanken die für unsere Forschungsfrage relevanten (und auch vorhandenen) Studien zuverlässig auffindet. Eine Barriere war dabei die unzureichende Kodierung methodischer Studien durch die Datenbankersteller. So überrascht es nicht, dass jede der nachfolgenden Untersuchungen (Concato, Benson, Ioannidis, Kunz 2002) weitere Übersichten eingeschlossen hat. Es besteht nach wie vor eine hohe Chance, dass eine erhebliche Anzahl an Übersichten, die unsere Einschlusskriterien erfüllen, nicht erfasst wurden. Allerdings ist weniger wahrscheinlich, dass qualitativ hochwertige bzw. rein methodisch ausgerichtete Studien^{129; 135-137}, die unsere Ergebnisse substantiell verändern würden, übersehen wurden.

Die Beschränkung auf eine deskriptive Analyse, mit z.T. semiquantitativen Vergleichen ist eine weitere Einschränkung unserer Untersuchung, die somit nur eine orientierende Abschätzung der Effekte ermöglicht. Allerdings haben die umfangreichen quantitativen Analysen in der Arbeit von Ioannidis¹⁴⁶ die Kernaussagen unserer Arbeit bestätigt.

2.1.1.5 Aktualisierung der Studie

Drei Jahre später haben wir nach erneuter Literatursuche (s. 2.3) und mit den im Intervall gesammelten Arbeiten die Evidenzlage unserer Übersicht nahezu verdoppelt. Nach dem Einschluss von 14 weiteren Übersichten und Kohorten mit mehr als 1700 Einzelstudien liegen jetzt 32 Übersichten mit mehr als 3000 Einzelstudien vor (Evidenztabelle Anhang 2, Kunz 2001¹⁴⁷), die nach der gleichen Methodik exzerpiert und analysiert worden sind. Die Aussagekraft der neuen Studien ist durch die methodischen Schwächen der meisten Arbeiten deutlich eingeschränkt. Auch die neu eingeschlossenen Übersichten fanden Effektverzerrungen, die in alle Richtungen gehen konnten, aber besonders bei den methodisch weniger validen Studien zu einer Überschätzung der Effektgröße führten. Diese Befunde bestätigen die Beobachtungen unserer ersten Arbeit und untermauern die getroffenen Schlussfolgerungen, dass 1) nicht (angemessen) randomisierte Studien über ein erhebliches Potential für verzerrte Einschätzungen des zugrunde liegenden Effekts verfügen, dass 2) ein offensichtlicher Trend zur Effektüberschätzung erkennbar ist, aber auch Unterschätzungen und Effekturnkehr auftreten können. Das nicht lösbare Problem besteht darin, dass die Richtung der Fehlschätzung ohne „Goldstandard“ nicht erkennbar und auch nicht korrigierbar ist. Somit ist es ein Paradox, dass die Unvorhersagbarkeit der Randomisierung den besten Schutz darstellt gegen die Unvorhersagbarkeit von Ausmaß und Richtung von Bias in Studien mit unzureichender Randomisierung. Wir haben dieses Phänomen als das „UNPREDICTABILITY PARADOX“ bezeichnet.

2.2 Weiterentwicklung der Diskussion durch Folgestudien

2.2.1 Zusammenfassung der Folgestudien

Die von uns aufgeworfene Diskussion wurde von anderen Arbeitsgruppen (Britton, Benson, Concato, Ioannidis) mit anderen methodischen Ansätzen fortgesetzt. Die einzelnen Arbeiten sind im Anhang 2 in Form von Evidenztabelle mit methodischen und inhaltlichen Details zusammengefasst. Im Folgenden werden die einzelnen Ergebnisse beschrieben und gegeneinander abgegrenzt.

Eine unserer eigenen Forschungsfragen „Erzeugen Nicht-RCTs systematisch größere oder kleinere Effektgrößen als RCTs?“ wurde in einem methodischen HTA-Bericht (UK) mit vier systematischen Übersichten und 14 Einzelvergleichen bearbeitet¹⁴⁸. Während die Ergebnisse ähnlich waren (Nicht-RCTs finden größere, kleinere oder identische Effekte wie RCTs ohne konsistente Richtung der Effektveränderung), betonten die Autoren ihre schlussfolgerung Befund, dass Nicht-RCTs nicht systematisch zu Effektüberschätzungen neigten, was eine längere Diskussion nach sich zog¹⁴⁸⁻¹⁵¹.

Zwei Publikationen im N Engl J Med (2000) präsentierten zwei ähnlich gelagerte Untersuchungen (Vergleich von randomisierten Studien mit (klassischen) Beobachtungsstudien): Benson kontrastierte mit 18 Einzelstudien („anecdotal evidence“) und drei Kohorten von Studien ausgewählte Beobachtungsstudien (retro-/prospektive Kohortenstudien, Querschnittsstudien, historische Kontrollen) mit thematisch gleichen RCTs¹⁵². Beim Vergleich der Lage der Punktschätzer der Beobachtungsstudie zum Konfidenzintervall des zugehörigen RCTs befanden sich die Punktschätzer meist innerhalb des Konfidenzbereichs. Die Autoren folgerten, dass Beobachtungsstudien und RCTs in der Regel vergleichbare Ergebnisse liefern und sich somit wichtige klinische Fragen auch durch Beobachtungsstudien (z.B. in Form von Datenbankanalysen) hochwertig und zuverlässig beantworten lassen. Die zweite Studie verglich den Effekt von Beobachtungsstudien (Fall-Kontroll- bzw. Kohortenstudien) mit dem von RCTs anhand von fünf definierten Interventionen¹²³, die in einem engen Sampling Frame (fünf Journals und einen Zeitraum von 1991 – 1995) identifiziert worden waren. In diesen Vergleichen lagen die Gesamteffekte der beiden Studientypen sehr nahe zusammen. Entsprechend lautete die Schlussfolgerung, dass RCTs und Beobachtungsstudien in der Regel zu

vergleichbaren Ergebnissen führen und (hochwertige) Beobachtungsstudien per se keine Überschätzung des Interventionseffekts bedingen.

In der jüngsten Publikation (JAMA 2001) untersuchte Ioannidis an 45 systematischen Übersichten (408 Primärstudien) die Frage: „Liefern RCTs und Beobachtungsstudien bei gleichen Behandlungen gleiche Ergebnisse?“¹⁴⁶. Dazu schloss er RCTs, Nicht-RCTs, Fall-Kontroll-, Kohorten- und Querschnittsstudien in die Untersuchung ein und führte eine quantitative Analyse mit Adjustierung für potentielle Confounder und Effektmodifikatoren durch. Trotz guter Korrelation zwischen den Behandlungseffekten der beiden Designkategorien (Korrelationskoeffizient 0.75) zeigten sich in Nicht-RCTs häufiger größere Effekte (25/45 versus 14/45 Topics). Sensitivitätsanalysen für unterschiedliche Definitionen von „Diskrepanz“ und alternative Analyseansätze (z.B. nach Designtyp) bestätigten die beobachteten Unterschiede in den Effektgrößen und den Trend zur Effektüberschätzung, aber auch das Vorkommen von Unterschätzungen durch Nicht-RCTs.

2.2.1.1 Kritische Analyse der gegenwärtigen Datenlage

Die Reaktion auf die „Unpredictability Paradox“-Studie und die Folgestudien belegen die Aktualität des Problems für Methodiker wie für Nutzer von Forschungsergebnissen. Im Folgenden sollen die wichtigsten Probleme und Grenzen der bisher durchgeführten Studien gemeinsam beleuchtet werden:

Verwendung von „Anecdotal Evidence“: Die Schlussfolgerungen von Britton¹⁴⁸ und Benson¹⁵² beziehen sich im Wesentlichen auf die Ergebnisse von „Fallserien“ einzelner Studien (Britton: 14/18 Vergleiche; Benson: 18/21 Vergleiche). Wie die einzelne klinische Studie unterliegt auch eine methodische Einzelstudie den gleichen Irrtumsmöglichkeiten durch Zufall, Bias oder Confounding, was der Aussagekraft und Übertragbarkeit Grenzen setzt. Da sich nahezu jeder methodische Aspekt mit einem effektvollen Beispiel unterstreichen oder widerlegen lässt, gilt eine methodische Beweisführung auf dieser Grundlage als überholt.

Methodische Qualität und klinische Heterogenität: Die Tatsache vergleichbarer Ergebnisse bedeutet nicht automatisch, dass die Primärstudien auch vergleichbare Fragestellungen, Populationen und Interventionen mit hohem methodischen Standard untersucht haben¹⁵³. In den meisten Arbeiten^{123;}

^{146; 148; 152} wurden die zugrundeliegenden Studien überhaupt nicht oder nur sehr oberflächlich auf

klinische und methodische Heterogenität untersucht. Die Überprüfung einiger der vorgenommenen Vergleiche^{123; 152} bestätigen die Berechtigung der Vorbehalte¹⁵⁴. So stammte z.B. die „RCT-Vergleichsgruppe“ aus Untergruppenanalysen von RCTs oder aus post hoc definierten Untergruppen¹⁵². In einem anderen Vergleich wurden randomisierte Hochdruckstudien mit Bevölkerungsdaten (ohne Angaben zur Intervention) kontrastiert¹²³. Eine umfassende Bewertung und transparente Darstellung von methodischer Qualität und klinischer Heterogenität und eine entsprechende Berücksichtigung dieser Ergebnisse bei der Gesamtbewertung war nur in den eigenen Arbeiten durchgeführt worden^{147; 151}.

Konfidenzintervalle als Analysegrundlage: Drei Untersuchungen^{123; 146; 152} verglichen die Lage des Punktschätzers aus der Beobachtungsstudie mit dem Konfidenzintervall des zugehörigen RCTs. Umschließt das Konfidenzintervall den Punktschätzer der Beobachtungsstudie, so wurde für beide Studientypen eine gemeinsame Grundgesamtheit angenommen. Allerdings verliert dieser Ansatz bei sehr breiten Konfidenzintervallen (kleine Stichprobe/große Variabilität) an Bedeutung, da der Punktschätzer der Beobachtungsstudie selbst dann noch innerhalb des Intervalls zu liegen kommt, wenn die Punktschätzer von RCT und Beobachtungsstudie weit auseinander liegen. Dieses Phänomen war besonders bei den Vergleichen von Benson¹⁵² und Ioannidis¹⁴⁶ offensichtlich, wobei Ioannidis das Problem thematisierte und das Ausmaß über eine Sensitivitätsanalyse zu erfassen suchte.

Wurde die Überlappung der Konfidenzintervalle „RCT/Beobachtungsstudie“ als „Gleichheitskriterium“ gewählt, zeigte sich ein ähnliches Problem. Je nach Wahl des zugrundeliegenden Meta-Analysemodells („Random effects model“ oder „Fixed effects model“) fand sich bei demselben Studienbeispiel (Mammographiescreening) eine Überlappung der Konfidenzintervalle der beiden Studientypen („Random effects model“)¹²³ oder nicht („Fixed effects model“)¹⁴⁶. Auch dieses Phänomen wurde in der zugehörigen Übersicht¹²³ nicht diskutiert.

Stichprobenrahmen: Eine entscheidende Einschränkung der Arbeiten von Concato und Benson ist das begrenzte Zeitfenster und die limitierte Auswahl an Datenbanken bzw. Journals, was zur Folge hatte, dass die Analysen nur auf einer kleinen Anzahl an Studien beruhten. Wichtige Meilensteinstudien¹⁵⁵⁻¹⁵⁷ und andere Studien¹⁵⁸, die die postulierten Effekte widerlegten, blieben dadurch unberücksichtigt.

2.2.1.2 Variabilität zwischen Studiendesigns: wünschenswert oder problematisch ?

Ist Variabilität innerhalb und zwischen den Designkategorien wünschenswert oder problematisch? Erstmals kann Variabilität ein Hinweis auf natürliche Vielfalt bei gleichen Einschlusskriterien und gleicher Umsetzung der Intervention in unterschiedlichen Settings bedeuten, wodurch es selbst bei hochwertigen Studien zu Unterschieden in wesentlichen Parametern kommen kann. Aber es müssen auch andere Ursachen für Variabilität bedacht werden: „Efficacy“-Studien, die die Wirksamkeit unter Idealbedingungen messen, können bei gleicher Intervention zu anderen Ergebnissen kommen als Effectiveness-Studien, die die Versorgungsrealität erfassen. Time Lag Bias kann sich bemerkbar machen, wenn nach frühen hochpositiven Effekten in Beobachtungsstudien bei weiteren Untersuchungen in RCTs mit größeren oder weniger selektierten Populationen ein „Verdünnungseffekt“ gefunden wird. Bei einem negativen Ergebnis können Designkategorien wie Beobachtungsstudien stärker von Publikationsbias betroffen werden als gut durchgeführte RCTs. Dies würde ebenfalls zu scheinbar größeren Effekten in Nicht-RCTs führen.

Daneben fallen noch andere wichtige Aspekte auf, die für ein besseres Verständnis der Datenlage näher untersucht werden müssten: Auf der Ebene der Primärforschung stellt sich die Frage, wie es überhaupt dazu kommt, dass die gleiche Forschungsfrage z.T. sogar zeitgleich mit zwei unterschiedlichen Designs untersucht wird. Auf der Ebene der Reviews ist unklar, weshalb in einigen Meta-Analysen überhaupt unterschiedliche Studientypen zur gleichen Fragestellung integriert werden¹²³ und in anderen nicht? Hier könnte ein neues Beispiel für Selektionsbias vorliegen, falls diese Kombination von Designs in der Existenz von Nicht-RCTs mit ähnlichen Ergebnissen wie den von RCTs begründet wäre.

2.2.2 Forschungsbedarf: Heterogenität - klinisch und methodisch

Die bisherigen Untersuchungen zeigen auf, dass die Schwierigkeiten, die einfache Frage nach der Bedeutung von Beobachtung oder Experiment valide zu beantworten, viel größer sind als ursprünglich angenommen und dass ein umfassender, systematischer Ansatz unter Einschluss aller existierender Studien dazu erforderlich ist. Ausserdem wurde deutlich, dass die klinische (Populationen,

Interventionen, Endpunkte) und methodische Heterogenität der Übersichten und möglicherweise sogar der Primärstudien herausgearbeitet und als Parameter in die Analyse eingeschlossen werden muss. Der dazu erforderliche Aufwand ist eine mögliche Erklärung, weshalb dieser Ansatz noch nicht umgesetzt wurde¹⁵⁹. Weitere vergleichende Untersuchungen sollten sich stärker auf die Unterschiede in der Effektgröße konzentrieren, an denen die Wirksamkeit einer Behandlung letztendlich gemessen wird.

2.2.3 Beobachtungsstudien – eine Standortbestimmung

Es besteht die Gefahr, dass durch die prominent publizierten Arbeiten von Concato und Benson die Beobachtungsstudie als Methode zur Bewertung der Wirksamkeit von Gesundheitsmaßnahmen (wieder) salonfähig wird. Die Aufweichung des Goldstandards RCT würde jedoch erhebliche Einbußen in der Qualität der Bewertung von Interventionen und einen Rückschlag in der Sicherstellung einer hochwertigen Medizin zum Schutze der Patienten bedeuten. Dabei besitzt die Beobachtungsstudie einen festen Platz als wichtige und informative Ergänzung zu RCTs, wenn es um die Erfassung der Wirksamkeit in der Routineversorgung, um späte Endpunkte oder unerwünschte Effekte geht¹⁶⁰⁻¹⁶² oder die Durchführung eines RCTs unethisch ist. Bei sehr ausgeprägten Effekten kann ein RCT überflüssig werden, bei sehr kleinen Interventionseffekten oder seltenen Ereignissen scheitert er schnell an der Durchführbarkeit. Hier wird man sich mit der größeren Unsicherheit über den wahren Behandlungseffekt arrangieren müssen, die der Einsatz von Beobachtungsstudien mit sich bringt.

Randomisierte kontrollierte Studien und Beobachtungsstudien haben bei der Bewertung von Interventionen derzeit klar zugeordnete Plätze. Empirische Methodikforschung muss zeigen, ob diese Zuteilung gerechtfertigt ist. Bevor sich unser Wissen darüber nicht weiter verdichtet hat, sollte der Status Quo, auch aus Respekt vor dem Patienten, nicht geändert werden.

2.3 Datenbank für methodische Studien – Grundstock für eine empirische Methodikforschung (eigene Untersuchungen)

Auch Empfehlungen, welche methodischen Verfahren zur Bewertung von Interventionen eingesetzt werden sollen, sollten möglichst auf guter Evidenz beruhen, und logisch plausible Argumente sollten einer empirischen Überprüfung standhalten.

1994 begann eine Gruppe von Wissenschaftlern (DL Sackett, AD Oxman, R Kunz, D Cowan, S Halperin) an der McMaster-University mit der Entwicklung eines Registers (Database of Trials and Reviews of Trials, TORTs) für Studien, die mit Daten aus der Patientenversorgung die Auswirkungen von methodischen Techniken auf die Effektschätzung untersuchten¹⁶³. Der erste Versuch (1995), auf der Grundlage der gesammelten Studien in elektronischen bibliographischen Datenbanken (MEDLINE) oder durch Handsearching eine systematische Suchstrategie zu entwickeln, scheiterte an der undifferenzierten Kodierung der Studien und den diffus verbreiteten Publikationsorten. Die unsystematische Sammlung über Referenzlisten, persönliche Kontakte und Vernetzung mit anderen Wissenschaftlern mit methodischem Interesse wurde fortgesetzt. Mit der Überführung der Datenbank in die Cochrane Collaboration als CRMD (Cochrane Review Methodology Database) wurde auch die Zielsetzung breiter (www.cochrane.de; Cochrane Methodology Review Group, „*Scope*“ and „*Topics*“). Angesichts des Eindrucks eines gestiegenen Interesses an empirisch-methodischer Forschung wurde von der Autorin während eines Forschungsaufenthalts 1999 an der Health Services Research Unit Folkehelse in Oslo, das Projekt, eine sensitive und effiziente systematische Suchstrategie zu entwickeln, nochmals aufgegriffen. Die Ergebnisse dieser Untersuchung sind bisher nicht veröffentlicht.

2.3.1 Fragestellung

Lässt sich für gängige medizinische Datenbanken eine effiziente Suchstrategie entwickeln, um Studien zu identifizieren, die die Auswirkung von definierten methodischen Verfahren zur Vermeidung von Bias auf Studienendpunkte untersuchen und dazu reale Patientendaten aus Studien zur

Gesundheitsversorgung verwenden. Relevante methodische Verfahren sind z.B. Randomisierung, Maskierung, Verblindung, qualitativ hochwertige verglichen mit weniger hochwertigen Studien, Vollständigkeit der Nachbeobachtung.

2.3.2 Methodik

2.3.2.1 Eingesetzte Datenbanken

Folgende Datenbanken wurden herangezogen: 1) MEDLINE (Ovid MEDLINE und Internet Grateful Med; Zeitfenster: 1996-7/1999) 2) Science Citation Index (SCI), eine Datenbank über die Zitierhäufigkeit von Artikeln 3) Die Cochrane Review Methodology Database (CRMD) der Cochrane Library, eine Bibliographie von methodischen Publikationen mit Relevanz für die Zusammenfassung von Interventionen aus der Gesundheitsversorgung (Stand 7/1999: ≥ 900 Zitate). Zum Zeitpunkt der Studie war die Entwicklung einer CRMD-eigenen Kodierung mit differenzierten methodischen Schlagwörtern weitgehend abgeschlossen Die CRMD wurde als Referenzdatenbank für die Suchstrategien verwendet.

2.3.2.2 Einschlusskriterien

Studienpopulation: Einzelstudien und Kohorten von Studien, Studien mit primärer methodischer Fragestellung, Studien zur Wirksamkeit von Interventionen in der Gesundheitsversorgung mit sekundär methodischer Fragestellung.


Interventionen: Vergleichende Untersuchungen über die Auswirkung von methodischen Verfahren auf Studienendpunkte mit Relevanz für die Durchführung von Reviews zur Vermeidung von Bias (beschrieben in dem Themenkatalog der Cochrane Empirical Methodological Studies Methods Group (Draft 30. Juni 1999; jetzige Version: www.cochrane.de; Cochrane Methodology Review Group, *Scope and Topics*).

Endpunkte: Qualitative oder quantitative Beschreibung der Auswirkung der methodischen Techniken.

2.3.2.3 Entwicklung der Suchstrategien

Folgende Strategien wurden exploriert: 1) Medline - MESH-Headings: Die Suche in Medline (OVID Medline) lief über die Medical Subject Headings (MeSH) - Indexierung, von denen für methodische Verfahren nur ein sehr limitiertes Set an Begriffen existiert. Diese wurden durch Subheadings wie *standard (st)*, *statistics & numerical (sn)* or *classification (cl)* weiter eingegrenzt und anschließend entsprechend der Fragestellung mit verschiedenen Varianten von „Random“ verknüpft. 2) Medline Feature „Related Articles“: In der seit 1998 existierenden „Related Articles“-Funktion von Internet Grateful Med (mit prädefinierten Algorithmen, die Studien auf der Grundlage von ähnlichen bzw. benachbarten Begriffen in Titel und Abstract suchen) wurden über (Index-) Publikationen weitere Artikel (www.ncbi.nlm.gov:80/entrez/query/static/computation.html) identifiziert und die unzureichende MeSH-Kodierung umgangen. Die Effizienz des Verfahrens wurde durch zwei Arten von Indexstudien ausgetestet: Studien mit primär methodischer Fragestellung^{127-129; 135; 137; 138; 151} und Studien mit primär klinischer und sekundär methodischer Fragestellung^{133; 134; 140; 143}. Studien, die bereits sicher in der CRMD-Datenbank eingeschlossen waren, wurden kein weiteres Mal berücksichtigt. 3) Im Science Citation Index wurden über die gleichen Indexstudien mit primär methodischer Fragestellung wie unter 2) weitere potentiell relevante Publikationen identifiziert. Die Vorgehensweise bei Screening und Auswahl der Studien, die durch zwei Wissenschaftler (RK und AO) erfolgte, sind für alle drei Suchansätze in Tabelle 6 wiedergegeben. Die Studien wurden in separaten Datenbanken gespeichert und die zusätzliche Ausbeute abgeglichen.

Tab. 6: Studienidentifizierung und Auswahlverfahren in den einzelnen Datenbanken

MeSH-Schlagwörter (Medline)	„Related Articles“-Funktion (Medline)	Science Citation Index
<ul style="list-style-type: none"> • Zuordnung von MeSH-Begriffen zu den für die Studienfrage relevanten CRMD-Schlagwörter • Suche mit MeSH-Begriffen (Tab.7a) • Kombination der Suchbegriffe mit verschiedenen Strategien für „Random“ (Tab.7b) 	<ul style="list-style-type: none"> • 7 methodische Schlüsselstudien • Suche in “Related Articles” <div style="text-align: center; margin: 10px 0;">  </div> <ul style="list-style-type: none"> • Klassifikation (R.K.) in: <ul style="list-style-type: none"> - (potentiell) geeignet - Quelle für weitere Zitate - ungeeignet • Überprüfung der (potentiell) geeigneten Studien durch 2. Wissenschaftler (A.O.) → Gesamtergebnis • Abgleich mit CRMD 	<ul style="list-style-type: none"> • 7 methodische Schlüsselstudien • Suche in “Science Citation Index”

CRMD: Cochrane Review Methodology Database

2.3.3 Ergebnisse

2.3.3.1 Ausbeute der MEDLINE-Suche

Die Suche über MeSH-Begriffe ist exemplarisch in Tab. 7a / 7b zusammengefasst. Tabelle 7a beschreibt die Trefferhäufigkeit einzelner MeSH-Begriffe ohne / mit Subheadings und Textwörter. Durch Kombinationen von methodischen Begriffen (immer mit Variationen des Begriffs „random“) konnten die sehr hohen Trefferzahlen deutlich reduziert werden (Tab. 7b). Von den 1594 initial identifizierten Zitaten wurden schließlich 49 Zitate als geeignet ausgewählt (Ausbeute: 3%). Mit den sieben methodischen Indexartikeln^{127-129; 135; 137; 138; 151} wurden über die “RELATED ARTICLES” Funktion insgesamt 863 Zitate identifiziert. Davon wurden 25 weitere, von den MeSH-Terms nicht entdeckte Studien eingeschlossen ($25/863 = 2,9\%$). Über die klinischen Indexstudien^{133; 134; 140; 143} fanden sich keine weiteren relevanten Studien.

Tabelle 8 vergleicht die Ausbeute der beiden Medlinestrategien mit den in der CRMD-Datenbank bereits vorhandenen Studien (inkompletter Referenzstandard). Die Ausbeute durch beide MEDLINE Strategien lag bei 3% (74 / 2455) aller gescreenten Studien. Davon befanden sich bereits 43 Studien (58%) in der CRMD, 31 neue Studien (42 %) konnten in die CRMD zusätzlich eingebracht werden (Zugewinn: 26%).

Tab. 7 a : Häufigkeit von MeSH-Begriffen und Textwörtern in Ovid-MEDLINE (1996 - 1999), die den relevanten Schlagwörtern der CRMD-Kodierung am ehesten entsprachen. Über methodische Subheadings (s.u.) wurden Studien mit methodischem Schwerpunkt herausgefiltert.

MeSH-headings / Textwörter	Treffer Ovid MEDLINE (1996 – 1999)	MeSH-headings / Textwörter	Treffer Ovid MEDLINE (1996 – 1999)
1. Bias (epidemiology) exp	7.467	7. Random allocation	5.940
2. Case control studies	66.715	8. Randomised controlled trials	11.008
3. Clinical trials	19.805	9. Randomised controlled trials / cl, mt, sn, st, td, ut	1348
4. Clinical trials / cl, mt, sn, st, td, ut	2.487	10. Research	82.960
5. Double blind method	15.108	11. Research / cl, mt, sn, td	3.128
6. Follow-up studies	55.704	12. Selection bias	435

Tab. 7 b: Verknüpfung methodischer MeSH-Begriffe und Textwörter mit Kombinationen des Begriffs „Random“ in Ovid Medline (1996 - 1999)

MeSH-Begriffe oder Textwörter kombiniert durch den Boole'schen Operator "OR"	Kombiniert mit MeSH-Begriffen durch den Operator "AND"	(Potentiell) geeignete / identifizierte Abstracts
1. Random allocation "OR" RCTs /cl, mt, sn, st, td, ut	Double blind method	23/138
2. RCTs (exp)	Selection bias	5/19
3. Random allocation "OR" RCTs (exp)	Research (cl, mt, sn, st, td)	10/80
4. RCTs (exp)	Case control studies (exp)	36/236
5. RCTs (exp)	Follow-up studies	18/245
6. RCTs /mt, sn, st, ut,	Follow-up studies	5/23
7. Random allocation "OR" RCTs (exp)	Bias (Epidemiology)	50/237
8. Random allocation "OR" RCTs /cl, mt, sn, st, td, ut	Clinical Trials /cl, mt, sn, st, td, ut	147/614
(mit Überlappungen) → 49 geeignete Studien		294/1592

Abkürzungen: Subheadings: cl=classification, mt=methods, sn=statistics & numerical, st=standards, td=trends, ut=utilization; exp=explode

Tab. 8: Ausbeute an geeigneten Studien aus der MEDLINE-Suche (MeSH-Begriffe und „RELATED ARTICLES“) und Zugewinn zu den in der CRMD bereits vorhandenen Studien

CRMD	MeSH (Gesamtzahl: 1592)	"RELATED ARTICLES" (Gesamtzahl: 863)	Gesamt (2455)
Bereits vorhanden	29	14	43
Neu eingeschlossen	20	11	31
	49	25	74

2.3.3.2 Ausbeute der Suche im Science Citation Index

Bei der Suche im SCI blieben von den 896 identifizierten Zitaten nach Entfernung von Duplikaten 56 geeignete Studien übrig (6,25%; Tab. 9). Dabei ergab der Indexartikel von Ken Schulz¹³⁷ über die Bedeutung der verdeckten Randomisierung die größte Ausbeute. Von den 56 identifizierten Studien wurden 21 (37%) neu in die CRMD aufgenommen, die restlichen Studien waren in der Datenbank bereits erfasst.

Tab. 9: Suche im Science Citation Index auf der Grundlage von sieben methodischen Indexstudien

	Jahre seit Publikation	Anzahl gescreenter Abstracts	für CRMD geeignete Studien	Geeignete Studien / Jahr
Kunz 1998 ¹⁵¹	2 J.	9	2	1/J.
Schulz 1995 ¹³⁷	5 J.	418	28	5,6 / J.
Ottenbacher 1992 ¹³⁸	8 J.	7	1	0.13 / J.
Emerson 1990 ¹²⁹	10 J.	74	10	1 / J.
Colditz 1989 ¹³⁵	10 J.	70	15	1.5 / J.
Sacks 1982 ¹²⁸	10 J.	161	10	1 / J.
Chalmers 1977 ¹²⁷	10 J.	157	6	0.6 / J.
Gesamte Ausbeute		896	56 *	

* ohne Duplikate

2.3.3.3 Sekundäre Fragestellungen

Methodikforschung im Trend

Eine Auflistung der eingeschlossenen Publikationen nach Publikationsjahr (Tab. 10) zeigt einen klaren Trend für eine Zunahme empirischer Untersuchungen innerhalb der Methodikforschung. Während zu Beginn der 90er Jahre (90/91) gerade sechs Studien in zwei Jahren publiziert worden waren, erfüllten in den Jahren 97/98 100 Studien die Einschlusskriterien.

Potential für Handsearching?

In unserer explorativen Studie von 1995 war ein punktuelles Handsearching in potentiell relevanten Journals (Contr Clin Trials, J Chron Dis, später J Clin Epidemiol; Stats Med) wenig ergiebig gewesen. Eine Auflistung der Medline-Suchergebnisse nach Publikationsorgan belegt erneut, dass Methodikstudien selten, aber in dem gesamten Spektrum der existierenden Journals publiziert werden (Tab. 11). 40 Journals veröffentlichten in dem Beobachtungszeitraum gerade eine Studie.

In unserer Stichprobe hatten nur das BMJ, Controlled Clinical Trials und JAMA mehr als fünf Studien

Tab. 10: Zeittrend für Publikationen empirischer Methodikstudien („geeignete“ und „potentiell geeignete“ Studien mit den beschriebenen Suchstrategien)

	MEDLINE	SCI
1999	4 + X #	10
1998	38	13
1997	35	14
1996	27	12
1995	10	7
1994	5	3
1993-92	3	7
1991-90	3	3
89 – 82	10	3

Zum Zeitpunkt der Suche war vermutlich nur ein Bruchteil der Publikationen von 1999 bereits in MEDLINE gelistet

Tab. 11: Potential für Handsearching?

Publikationsorgane der 74 relevanten Studien aus der kombinierten MEDLINE –Suche

Zeitschrift	Anzahl Publikationen
BMJ	8
Controlled Clinical Trials	7
JAMA	6
Lancet	4
Archives of Internal Medicine	2
Journal of Family Practice	2
Epidemiologische und statistische Zeitschriften (n=3)	5
40 Journals (je 1 Publikation)	40

publiziert. Systematisches Handsearching erscheint auf der Grundlage dieser Daten unverändert unergiebig.

Wissenschaftliche Wertschätzung von Methodikstudien

Tabelle 12 listet die Impactfaktoren der Zeitschriften, in denen die Studien von Tabelle 11 publiziert wurden. Die überwiegende Mehrzahl der Studien wurde nicht in den theoretisch-methodischen, sondern in medizinischen Journals mit hohen Impactfaktoren publiziert. Danach finden Methodikstudien Eingang und wissenschaftliche Anerkennung in hochwertigen Journals.

Tab. 12: Publikationsorgan nach Fachgebiet und Impactfaktor der Zeitschrift

Journal nach Fachgebiet	Anzahl der Zeitschriften	Impact factor (1999) (median)	Impact factor (1999) (mean)
Spezialisierte Journal	27	2.49	2.33
Allgemein medizinische Journal	8	5.92*	6.03°
Public Health und epidemiologische Journal	7	1.69	1.42
Methodik- und Statistik-Journal	5	1.34	0.93
Allgemeinmedizinische Journal	2	1.70	1.70
Insgesamt	49	1.91*	2.76°

*ohne N Engl J Med; sonst: IF: 6.70, bzw. Gesamtbewertung: IF: 1.97

° ohne N Engl J Med; sonst: IF: 8.56, bzw. Gesamtbewertung: IF: 3.32

2.3.4 Diskussion

Trotz der fokussierten primären Fragestellung (empirische Vergleiche über die Auswirkung von methodischen Verfahren auf Studienendpunkte) ist es nicht gelungen, in den elektronischen bibliographischen Datenbanken Medline und Science Citation Index eine effiziente systematische Suchstrategie zu entwickeln, die routinemäßig einsetzbar wäre, um mit ausreichender Sicherheit und Ergiebigkeit die relevanten Publikationen zu erfassen. Verglichen mit den Suchstrategien von 1994/95 hat sich die Ausbeute insbesondere im „Science Citation Index“ und in der „Related Articles“-

Variante von MEDLINE deutlich verbessert, was sich durch ein gestiegenes Interesse an methodischen Fragestellungen, die zunehmende wissenschaftliche Anerkennung und gute Publizierbarkeit in renommierten Zeitschriften erklären lässt.

Die explorative Studie bestätigt die Notwendigkeit eines Registers wie dem CRMD als eine Ressource, mit der durch spezifischere Indexierung der relevanten Arbeiten und unter Einführung neuer Schlagwörter die Voraussetzungen für weitere empirische Methodikforschung geschaffen werden. Angesichts des nachgewiesenen Trends sollte in zwei bis drei Jahren ein erneuter Versuch zur Entwicklung einer geeigneten Suchstrategie unternommen werden. Bis dahin müssen die Grundlagen für empirische Studien weiterhin mühsam und unvollständig zusammengetragen werden, was die Schwierigkeit der Durchführung solcher Studien erhöht und die Zuverlässigkeit der Studienergebnisse limitiert.

3 Risiko und Wirksamkeit, Behandlungsschwelle und Bias: Potential für klinische Fehlentscheidungen

Aufbau des Kapitels

Im vorherigen Kapitel wurde nachgewiesen, wie wichtig die Integration von methodischen Techniken in das Studiendesign ist, um zuverlässig gegen Bias zu schützen und den „wahren“ Effekt einer Intervention zu erfassen. In dem Kapitel „Risiko und Wirksamkeit, Behandlungsschwelle und Bias“ wird diese Erkenntnis auf konkrete Versorgungssituationen angewandt. Die quantitative Beschreibung von Risiko und Wirksamkeit (mit klinisch-epidemiologischen Grundbegriffen) kann helfen, in der Versorgung individueller Patienten differenzierte Entscheidungen zu fällen. Im ersten Abschnitt (3.1) wird der Nutzen klinisch-epidemiologischer Grundlagen für die ärztliche Entscheidungsfindung diskutiert und an mehreren Studien beispielhaft das (Fehl-)Verständnis dieser Konzepte unter Ärzten und anderen Gesundheitsberufen aufgezeigt. Diese Grundbegriffe werden in unserer Simulationsstudie (3.2) wieder aufgegriffen, in der wir für verschiedene Annahmen von Patientenrisiko, Wirksamkeit der Behandlung und Bias im Studiendesign demonstrieren, wie es gerade bei den häufig auftretenden Niedrigrisikokonstellationen zu Fehlentscheidungen kommen kann. Im Rahmen des Clearingverfahrens Hypertonie (3.3) kommen diese Begriffe erneut zum Einsatz. Nach einem Clearingprozess von deutschen und internationalen Hypertonieleitlinien nach EbM-Kriterien haben wir in den Empfehlungen für eine nationale Hypertonieleitlinie einen differenzierten Umgang mit der Interaktion von zugrundeliegendem Patientenrisiko und Behandlungseffekt als besonders wünschenswert hervorgehoben. Die Umsetzung der Empfehlungen wird diskutiert.

Eigene eingeschlossene Publikationen zum Thema:

- Kunz R, Neumayer HH, Khan KS. When small degrees of bias in randomized trials can mislead clinical decisions: an example of individualizing preventive treatment of upper gastrointestinal bleeding. *Crit Care Med* 2002;30:1503-7.
- Ärztliche Zentralstelle Qualitätssicherung. Leitlinien-Clearing-Bericht "Hypertonie". Autoren: Donner-Banzhoff N, Echterhoff HH, Hense HW, Kunz R, Sawicki P, Thurmann P *et al.* Germering/München: W.Zuckschwedt Verlag, 2000.
- Donner-Banzhoff N, Echterhoff HH, Hense HW, Kunz R, Sawicki P, Thurmann P *et al.* Guidelines Clearing House Statement "Hypertension". Summary and recommendations for a rational hypertension guideline in Germany. *Z.Arztl.Fortbild.Qualitatssich.* 2000;94:341-9.

- Kunz R, Vist G, Oxman A. D. Randomisation to protect against selection bias in healthcare trials. The Cochrane Library (3). 2002.
- Khan KS, Kunz R, Kleijnen J, Antes G. Systematic Reviews to Support Evidence-based Medicine: How to review and apply findings of health care literature. London: The Royal Society of Medicine Press, 2002.
- Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.

3.1 Die klinische Epidemiologie in der ärztlichen Entscheidungsfindung

3.1.1 Wirksamkeit von Interventionen

Um beim Management von Patienten differenzierte Entscheidungen zu fällen, benötigt ein Arzt spezifische Informationen über die beabsichtigte Behandlung und den betroffenen Patienten, die er gegeneinander abwägen muss: Was sind die Konsequenzen, nichts zu tun (Ausgangsrisiko des Patienten) verglichen mit dem potentiellen Nutzen der Behandlung (Ausmaß der Risikosenkung)? Wie wahrscheinlich und wie groß sind die potentiellen Nebenwirkungen durch die Behandlung? Welches sind die Patienten mit hohem Risiko und / oder guter Ansprechbarkeit auf die Behandlung? Wie wirksam ist die vorgesehene Behandlung verglichen mit anderen präventiven oder therapeutischen Maßnahmen? Dafür sind einheitliche, aussagekräftige und verständliche Maßzahlen erforderlich, die möglichst viele der aufgeführten Aspekte abbilden. Für binäre Ereignissen stehen dafür das relative Risiko (RR), die relative Risikoreduktion (RRR) und die Odds Ratio (OR), die absolute Risikoreduktion (ARR) und die Number Needed to Treat (NNT) zur Verfügung.

3.1.1.1 Relative Maßzahlen

Das relative Risiko (RR) ist das Verhältnis zwischen dem Risiko für ein negatives Ereignis (z.B. Tod) in der Behandlungsgruppe bezogen auf das Risiko für ein negatives Ereignis in der Kontrollgruppe. Entsprechend beschreibt die relative Risikoreduktion (RRR) die Differenz der Ereignisraten zwischen Kontroll- und Behandlungsgruppe, dividiert durch die Ereignisrate in der Kontrollgruppe (Tab. 3.1). Die RRR wird in der Umgangssprache intuitiv bei der Beschreibung von Risiken verwendet und somit auch häufig in der Kommunikation zwischen Ärzten und Patienten eingesetzt.

Tab. 3.1: Relatives Risiko und Odds Ratio: Identische Studienergebnisse, ausgedrückt als RR, RRR und OR am Beispiel der Ulkusprophylaxe durch H₂-Blocker¹⁶⁴ bei unterschiedlichem Patientenrisiko (hoch: 30%; gering: 3%).

Risiko	Ausgangsrisiko (Ulkus)		Odds	Ausgangsrisiko (Ulkus)	
	Hoch	gering		hoch	gering
Risiko _{Kontr} (n=100)	$\frac{30}{100}=0.3$	$\frac{3}{100}=0.03$	Odds _{Kontr} (n=100)	$\frac{30}{70} = 0.43$	$\frac{3}{97} = 0.03$
Risiko _{H2-Blocker} (n=110)	$\frac{20}{110}=0.18$	$\frac{2}{110}=0.018$	Odds _{H2-Blocker} (n=100)	$\frac{20}{90} = 0.22$	$\frac{2}{108} = 0.0185$
RR	$\frac{0.18}{0.30}=0.6=\frac{0.018}{0.03}$		OR	$\frac{0.22}{0.43} = 0.52$	$\frac{0.0185}{0.03} = 0.62$
RRR	1-0.6 = 0.4 = 40%				

RR=relatives Risiko; RRR=relative Risikoreduktion; OR=Odds Ratio

Die „Odds“ für ein Ereignis ist das Verhältnis der Wahrscheinlichkeit, dass das Ereignis stattfindet zur Wahrscheinlichkeit, dass es nicht stattfindet. Die Odds Ratio (OR) ist demnach das Verhältnis

$\frac{Odds_{Intervention}}{Odds_{Kontrolle}}$. Während die OR wegen ihrer stabilen statistischen Eigenschaften von Methodikern geschätzt wird, fehlt Ärzten und Patienten in der Regel der Zugang zu dieser Maßzahl¹⁶⁵.

Da relative Maßzahlen Grundrisiko und Behandlungsnutzen für den Patienten nicht abbilden, erscheint bei klinischen Entscheidungen die Wirksamkeit einer Behandlung bei hohem Ausgangsrisiko ungerechtfertigt pessimistisch, bei niedrigem Ausgangsrisiko überoptimistisch.

3.1.1.2 Absolute Maßzahlen

Aussagekräftigere Informationen über die reale Wirksamkeit einer Maßnahme liefert die absolute Risikoreduktion (ARR), die die Differenz zwischen den Ereignissen von Kontroll- und Behandlungsgruppe beschreibt und dabei das Grundrisiko des Patienten einbezieht. Dennoch hat sie keinen Eingang in die ärztliche Kommunikation gefunden. Erst in der reziproken Darstellung als „Number Needed to Treat“ (Anzahl an Patienten, die über einen bestimmten Zeitraum behandelt werden müssen, um ein zusätzliches negatives Ereignis zu verhindern¹⁶⁶) wurde die Information in ein leicht verständliches Format gebracht.

Tab. 3.2:
Number
Needed to
Treat

Interaktion
zwischen
Ausgangsrisiko
und
Wirksamkeit
einer
Intervention.

	A) Ausgangsrisiko hoch (80%) Wirksamkeit mäßig (RRR: 15%)	B) Ausgangsrisiko mäßig (30%) Wirksamkeit gut (RRR: 40%)	C) Ausgangsrisiko* gering¹⁶⁷ (1,7%) Wirksamkeit gut (RRR: 33%)
Risiko_{Kontr}	$\frac{80}{100} = 0.8$	$\frac{30}{100} = 0.3$	$\frac{52}{3293} = 0.017$
Risiko_{H2-Blocker}	$\frac{68}{100} = 0.68$	$\frac{18}{100} = 0.18$	$\frac{38}{3302} = 0.012$
ARR	$0.8 - 0.68 = 0.12$	$0.3 - 0.18 = 0.12$	$0.017 - 0.012 = 0.005$
1/ARR = NNT	$\frac{1}{0.12} = 8,3Pat$	$\frac{1}{0.12} = 8,3Pat$	$\frac{1}{0.005} = 200Pat_{5Jahre}$

* Ausgangsrisiko für Tod durch KHK

Durch die Wechselwirkung zwischen Ausgangsrisiko und Wirksamkeit einer Intervention profitieren Patienten mit hohem Risiko (Beispiel A: 80%) von weniger wirksamen Maßnahmen (RRR 15%) in dem gleichen Ausmaß wie Patienten mit geringerem Grundrisiko (Beispiel B: 30%), die Maßnahmen von größerer Wirksamkeit (RRR 40%) erhalten (Tab. 3.2). In beiden Beispielen resultiert eine NNT von acht Patienten. Dagegen führt selbst eine effektive Behandlung (RRR 33%) nur zu einem bescheidenen Behandlungsnutzen (NT_{5 J.}: 200), wenn das Ausgangsrisiko gering ist (Beispiel C: 1,7%). Der Zusammenhang von Ausgangsrisiko und NNTs bei konstantem Behandlungseffekt lässt sich grafisch in der sogenannten NNT-Kurve darstellen (s. auch 3.2.2.2 und Abb. 3.1). Verglichen mit den relativen Maßzahlen bieten die NNTs einen besseren Schutz gegen Fehleinschätzungen über den Nutzen von Maßnahmen.

3.1.1.3 Übertragbarkeit auf den Versorgungsalltag: Grenzen und Beschränkungen der Maßzahlen

Von allen Maßzahlen erfüllen die NNTs am ehesten die in 3.1.1. beschriebenen Anforderungen für den klinischen Einsatz. Dennoch haben auch NNTs ihre Grenzen, insbesondere bei der Anwendung auf andere Populationen oder Individuen¹⁶⁸: Durch Selektion des Patientenspektrums, intensive Unterstützung bei der Compliance und andere studienspezifische Maßnahmen¹ wird selbst die aus „Effectiveness“-Studien geschätzte Wirksamkeit einer Behandlung für die Routineversorgung

¹ So wurden z.B. für die WOSCOP-Studie³⁰⁸, die den ersten Nachweis einer erfolgreichen Primärprävention von KHK durch Lipidsenkung lieferte, 80.000 Personen gescreent, 20.000 in die engere Auswahl eingeschlossen und 6.000 schliesslich rekrutiert. Zur Steigerung der Compliance fand alle 3 Monate eine Nachbeobachtung statt. Die Studie selbst wurde in Schottland, einer der Gegenden Europas mit der höchsten KHK-Prävalenz, durchgeführt.

höchstwahrscheinlich überschätzt. Um Unterschiede in den Ausgangsrisiken zwischen dem eigenen Patienten und dem durchschnittlichen Studienpatienten anzupassen, wurde aus dem Verhältnis $\frac{\text{Ausgangsrisiko eigener Patient}}{\text{Ausgangsrisiko typ. Studienpatient}}$ der Korrekturfaktor „f“ entwickelt¹⁶⁹. Die Division der Studien-NNTs durch den Korrekturfaktor „f“ ergibt somit einen Näherungswert für die NNT des eigenen Patienten. Bei geringerem Patientenausgangsrisiko würden die in der Studie ermittelten NNTs gegebenenfalls deutlich ansteigen und den Behandlungsnutzen erheblich relativieren und umgekehrt. Aus klinischer Sicht besteht für die Ärzte die eigentliche Herausforderung in einer angemessenen Einschätzung des patientenspezifischen Grundrisikos bei den unterschiedlichen Krankheiten. Trotzdem bedeuten die NNTs für Ärzte und Patienten eine erhebliche Bereicherung für die Interpretation von Studienergebnissen oder die Kommunikation über den Nutzen (bzw. fehlenden Nutzen) von Maßnahmen.

Die Argumente für die besondere Eignung der NNTs als Kommunikationsmittel werden durch neuere Erkenntnisse aus der Kognitionspsychologie¹⁷⁰⁻¹⁷² unterstützt. Experimente zeigten, dass die Darstellung von Risiken in Wahrscheinlichkeiten (z.B. 0.2 oder 20%) selbst bei Akademikern, die berufsmäßig mit dem Risikobegriff umgehen (z.B. Ärzte oder Richter), zu erheblichen Fehlinterpretationen der Aussagen führte, während die Darstellung der gleichen Situationen in natürlichen Häufigkeiten (z.B. 1 von 5) korrekt interpretiert wurde¹⁷³. Verschiedene Studien zu medizinischen Entscheidungssituationen haben diese Interpretationsprobleme („Framingeffekt“) wiederholt auch bei Ärzten nachgewiesen, die in Abhängigkeit von der Darstellungsform der gleichen Information schlussendlich zu diskrepanten Behandlungsentscheidungen kamen¹⁷⁴⁻¹⁷⁸.

3.1.2 Nutzen für den Patienten: Ausgangsrisiko und Risikoreduktion

Die Wirksamkeit einer Intervention und das patientenspezifische Ausgangsrisiko bestimmen wesentlich den Nutzen der Therapie. Die Prädisposition für negative Ereignisse kann dabei jedoch stark variieren. Patienten nach ihrem Risiko (d.h. ihrer Prognose) zu stratifizieren und sie einer angemessenen Behandlung zuzuführen, gehört zu den Grundfertigkeiten der ärztlichen Profession. Doch wie gut beherrschen Ärzte diese Fertigkeiten?

3.1.2.1 Wahrnehmung von Ausgangsrisiko und Behandlungsnutzen – Drei Beispiele

In einer amerikanischen Studie sollten Hausärzte und Kardiologen für typische Szenarien aus der hausärztlichen Praxis mit einfachen Risikokonstellationen (Mann mittleren Alters mit isolierter Hypercholesterinämie; ältere Person mit stabile A. pectoris bei Hauptstammstenose) die 5- bzw. 3-Jahresprognose für ein kardiales Ereignis sowie den relativen und absoluten Behandlungsnutzen von Cholesterinsenkung bzw. einer Bypassoperation schätzen¹⁷⁹. Nach Studienlage lag das 5-Jahres-Risiko für Myokardinfarkt in dem Hypercholesterinämie-Beispiel bei 6% (alle Angaben als Median), das 3-Jahres-Risiko für Überleben bei 70%. Die Hausärzte überschätzten das Risiko um mehr als 300% (Grundrisiko: 20%), die Kardiologen immerhin um fast 70% (Grundrisiko: 10%). Das Risiko der Hauptstammstenose wurde dagegen annähernd korrekt eingeschätzt.

Bei der Bewertung des Behandlungsnutzens fielen die Fehlentscheidungen nicht weniger ausgeprägt aus: Bei der Hypercholesterinämie überschätzten die Hausärzte die RRR um mehr als 120% ($RRR_{\text{Ärzte}}: 41\%$ verglichen mit $RRR_{\text{Studien}}: 18\%$), bei der ARR um 800% ($ARR_{\text{Ärzte}}: 8\%$ verglichen mit $ARR_{\text{Studien}}: 1\%$). Bei der Hauptstammstenose wurde die Wirksamkeit dagegen erheblich unterschätzt ($RRR_{\text{Ärzte}}: 6\%$ verglichen mit $RRR_{\text{Studie}}: 29\%$; $ARR_{\text{Ärzte}}: 20\%$ verglichen mit $ARR_{\text{Studie}}: 20\%$). Die Einschätzungen der Kardiologen waren insgesamt realistischer, überschätzten aber insbesondere die Wirksamkeit der primärpräventiven Maßnahme noch immer erheblich.

Wie ist unter Ärzten das Verständnis für Risiko und Behandlungsnutzen in komplexeren Versorgungssituationen, wenn Patienten multiple Risikofaktoren und Begleiterkrankungen haben, die über verschiedene Behandlungsansätze modifiziert werden können? In einer kanadischen Studie¹⁸⁰ überschätzten Allgemeinärzte das kardiale 8-Jahres-Risiko eines typischen Durchschnittskanadiers (für A. pectoris, Myokardinfarkt oder koronaren Tod) um den Faktor 6,7 verglichen mit der Studienlage (für Kanada), das 8-Jahres-Risiko eines 40jährigen Mannes mit komplexem KHK-Risiko-Profil um den Faktor 4,4. Beim Vergleich des relativen Risikos des KHK-Patienten zu dem des Durchschnittskanadiers entsprachen die Schätzungen der Ärzte ziemlich genau der Studienlage. Auch bei den Erwartungen an den Behandlungserfolg gemessen an gewonnenen Lebensjahren hatten die Ärzte sowohl über die Wirkung der Einzelmaßnahmen (Raucherentwöhnung, Blutdruck- und Cholesterinsenkung) wie auch über die des Maßnahmenpaketes im Prinzip keine Vorstellungen über

den absoluten Nutzen der Maßnahmen, sondern überschätzten den zu erwartenden Gewinn an Lebensjahren um ein Vielfaches.

Fehleinschätzungen in der Größenordnung des Nutzens werden für die Versorgung jedoch nur dann bedeutsam, wenn dadurch Behandlungsentscheidungen fehlgeleitet werden. Um diese Frage zu untersuchen, sollten Allgemeinärzte für fünf Patienten mit unterschiedlichem Risikoprofil das kardiale 5-Jahres-Risiko ohne Behandlung einschätzen und ihre Entscheidung für bzw. gegen eine Therapie explizit begründen¹⁸¹. Die Bewertungen wurden mit großen Studien und nationalen Therapieempfehlungen verglichen. Durch die Wahl des Blutdruckwertes (statt dem Gesamtrisiko) als häufigste Behandlungsindikation kam es besonders bei extremen Risikokonstellationen zu Fehlentscheidungen: Ältere multimorbide Patienten wurden in ihrem Risiko unterschätzt und unterbehandelt, jüngere Patienten wurden in ihrem Risiko überschätzt und unangemessen medikamentös eingestellt. Die Behandlung der Patienten unabhängig vom Gesamtrisiko hatte in der Studie erhebliche Fehlallokationen der Ressourcen zur Folge.

3.1.2.2 Konsequenzen für die Versorgung

Ärzte haben selbst bei gängigen Problemen Schwierigkeiten, das Ausgangsrisiko ihrer Patienten adäquat zu bewerten¹⁸². Während sie über ein vergleichsweise solides Verständnis für den relativen Behandlungsnutzen verfügen, können sie Patienten mit hohem Risiko und größtem absoluten Behandlungsnutzen nur unzureichend von solchen mit niedrigem Risiko und geringem absoluten Nutzen abgrenzen, was unvermeidlich zu Fehlallokationen der vorhandenen knappen Ressourcen führt. So gibt es besonders aus dem primärpräventiven Bereich Hinweise, dass Ärzte bestimmte, vielfach propagierte Maßnahmen deutlich seltener verschreiben würden, wenn ihnen der geringe absolute Nutzen voll bewusst wäre^{174; 175; 179}. Angesichts der laufenden gesundheitspolitischen Diskussionen um Zweckmäßigkeit und Angemessenheit von Behandlungen müsste dieses Thema intensiver beforscht und das Bewusstsein für den Zusammenhang von Risiko und Wirksamkeit unter den praktizierenden Ärzten deutlich geschärft werden. Bei der Formulierung von Behandlungsempfehlungen sollte es einen hohen Stellenwert erhalten.

3.1.3 Behandlungsschwellen und ihre Determinanten

3.1.3.1 Wo liegen die Behandlungsschwellen? Eine empirische Momentaufnahme

Damit eine Behandlung begonnen wird, muss der Nutzen, d.h. der kombinierte Effekt aus Ausgangsrisiko und Risikoreduktion eine (imaginäre) Schwelle überschreiten. Jeder Arzt hat bei seinen Entscheidungen implizit eine Therapieschwelle im Kopf. Doch wo ist die Schwelle angesiedelt? Wie groß muss der Nutzen sein, um eine Behandlung zu rechtfertigen? Selbst das Sozialgesetzbuch V bleibt bei der Definition, was „angemessen“ und „zweckmäßig“ bedeutet, vage. Die wenigen Studien zu dieser Frage beschreiben unter Ärzten, Gesundheitsberufen, Betroffenen und gesunden Verbrauchern in den Auffassungen von "angemessen“ erhebliche Diskrepanzen: In der Studie über die Behandlung bei Hypercholesterinämie¹⁷⁹ fanden Allgemeinärzte erst eine $ARR_{\text{Myokardinfarkt}}$ von 8-10% für einen Therapiebeginn als angemessen, während Kardiologen die Behandlung bereits bei einer ARR von 3-4% empfahlen. Bei einer Befragung über persönliche Behandlungsschwellen für milde Hypertonie¹⁸³ forderten gesunde Personen aus der Allgemeinbevölkerung und Pflegekräfte von der Behandlung einen viel höheren Nutzen (NNT 33; Range: 12-50) als Allgemeinärzte (NNT: 50; Range: 33-100) oder gar Chefärzte (NNT: 100; Range: 50-250), bevor sie sich der Behandlung unterziehen wollten. Bei der Darstellung des Nutzens in den unterschiedlichen Maßzahlen zeigte sich bei Patienten und Verbrauchern der gleiche Framingeffekt wie bei Ärzten^{174-176; 184}. Wurde der Nutzen als RRR dargestellt, akzeptierten 92% der Patienten eine antihypertensive Behandlung, in der Darstellung als NNT waren es 68%, dargestellt als „persönlicher Nutzen“ sank die Zahl der Behandlungswilligen auf 44%. Auch kulturelle Einflüsse prägen Behandlungsschwellen: In einem internationalen Vergleich sollten europäische und US-amerikanische Onkologen entscheiden, ob sie zwei näher beschriebene Patientinnen mit Mammakarzinom und etwas unterschiedlichen Begleitumständen mit adjuvanter Chemotherapie behandeln würden. Studien hatten für die adjuvante Chemotherapie einen statistisch signifikanten Effekt nachgewiesen, der aber klinisch nur von marginaler Bedeutung war. Bei den US-amerikanischen Ärzten entschieden sich 93% (Fall 1) bzw. 85% (Fall 2) für eine Behandlung, bei den europäischen Ärzten waren es lediglich 68% bzw. 67%¹⁸⁵.

In diesen und anderen Beispielen hatten Spezialisten eine deutlich niedrigere Behandlungsschwelle als Allgemeinärzte oder Verbraucher. Möglicherweise beeinflusst das durchschnittliche Risiko der täglichen Arbeitsumgebung die Risikowahrnehmung der Ärzte und entsprechend auch ihre Behandlungsschwelle. Dennoch stellt sich erneut die Frage, wer nach welchen Kriterien die für eine Behandlung angemessene Schwelle festlegen soll.

3.1.3.2 Die Behandlungsschwelle beim individuellen Patienten – Grenzlinie zwischen Nutzen, Schaden und persönlichen Werten

Eine Einschränkung bei den oben aufgeführten Studien ist die Reduktion der Fragestellung auf den Nutzen der Behandlung. Andere wichtige Komponenten wie Häufigkeit und Schweregrad von Nebenwirkungen, Behandlungsalternativen oder benötigte Ressourcen (über die Bedeutung der „Kosten“ gehen die Vorstellungen der Ärzte weit auseinander) sind in die Überlegungen nicht eingeflossen. Diese Elemente finden in die ärztlichen Entscheidungsfindung in der Regel intuitiv und mit unterschiedlichen Gewichtungen Eingang. Die geringe Transparenz dieser Prozesse ist jedoch ein Grund für die erhebliche Praxisvariation bei ähnlich gelagerten Krankheitsproblemen mit eigentlich guter Datenlage, die u.a. in Berichten wie dem Gutachten des Sachverständigenrats für die Konzertierte Aktion im Gesundheitswesen wiederholt beschrieben wurden¹⁸⁶.

3.2 Behandlungsschwelle und Bias: Potential für klinische Fehlentscheidungen (eigene Untersuchungen)

Simulationsstudie zur Individualisierung von Behandlungen und zum Einfluss von Bias auf ärztliche Entscheidungen

In dem folgenden Beispiel werden die vorgestellten klinisch-epidemiologischen Grundbegriffe auf realen Patientenszenarien angewandt und ihr Einfluss auf die ärztliche Entscheidungsfindung untersucht. Bei den bisherigen Beispielen war man implizit immer davon ausgegangen, dass die Schätzungen der Wirksamkeit einer Behandlung zuverlässig und präzise sind. Im Kapitel 2 hatten wir jedoch ausführlich diskutiert^{147; 187}, wie vorsichtig man mit dieser Annahme umgehen muss. Dabei ist

Bias, eigentlich das Feld von Methodikern, Epidemiologen und Biometrikern, für den praktizierenden Arzt meist nur von marginalem Interesse. Die Situation ändert sich jedoch, wenn es durch die Existenz von Bias zu offensichtlichen Fehlentscheidungen in der Patientenversorgung kommt. Die relevanten Einflussfaktoren und Wechselwirkungen, unter denen das Risiko für Fehlentscheidungen besonders groß ist, besser zu verstehen, ist ein weiteres Ziel dieser Arbeit:

3.2.1 Fragestellung

Unsere Studienfrage lautet: Unter welchen Umständen besteht in der Versorgung individueller Patienten ein erhöhtes Risiko für ärztliche Fehlentscheidungen, wenn Behandlungseffekte verzerrt erfasst werden?

Dabei erfolgt die Bearbeitung der Frage in drei Schritten:

1. Individualisierung von Studienergebnissen für Therapieentscheidungen in der ärztlichen Praxis
2. Auswirkung von Ungewissheit in der Präzision der Schätzung des Therapieeffektes
3. Simulation über die Konsequenzen von Bias auf die Entscheidungsfindung bei unterschiedlichen Konstellationen:
 - a. Ausgeprägte Bias bei großem Behandlungseffekt
 - b. Ausgeprägte (bzw. mäßig große) Bias bei moderatem Behandlungseffekt

3.2.2 Methodik

3.2.2.1 Klinischer Rahmen des Modells

Patienten: **Modellpatient A** ist ein beatmeter, intensivpflichtiger Patient mit respiratorischer Insuffizienz und Gerinnungsstörungen; **Modellpatient B** ist ausschließlich wegen respiratorischer Insuffizienz intensivpflichtig und beatmet.

Zur Beurteilung des Ausgangsrisikos der Modellpatienten für eine gastrointestinale (GI-)Blutung wurden die Schätzungen aus einer großen multizentrischen Kohortenstudie mit mehr als 2000 Patienten herangezogen¹⁸⁸. Nur zwei von zehn untersuchten Risikofaktoren für klinisch relevante GI-Blutung zeigten sich nach multivariater Analyse statistisch signifikant: respiratorische Insuffizienz und Gerinnungsstörungen. Demnach beträgt für den Hochrisikopatienten A das Risiko für eine klinisch

bedeutsame GI-Blutung während des Aufenthaltes auf der Intensivstation 8,4%, für den Niedrigrisikopatienten B lediglich 2%.

Intervention und Endpunkt: H₂-Blocker haben sich als hochwirksame Medikamente bewährt, um Stressulzera, eine typische Komplikation intensivpflichtiger Patienten zu vermeiden. In einer hochwertigen Meta-Analyse¹⁶⁴ war im Vergleich zu keiner Behandlung ein Prophylaxeeffekt für GI-Blutungen von OR_{Gesamt} 0.58 nachgewiesen worden.

3.2.2.2 Methodische Grundbegriffe des Modells

- Die **Number Needed to Treat (NNT)** wurde bereits unter 3.1.1.2. abgeleitet.
- Die **NNT-Kurve**. Die Wechselwirkung zwischen Ausgangsrisiko und Behandlungseffekt lässt sich grafisch als NNT-Kurve darstellen. Dabei werden die NNTs einer Intervention über unterschiedliche Patientenrisiken aufgetragen. Die zugrundeliegende Annahme eines konstanten Behandlungseffekts über unterschiedlichen Risiken wurde kürzlich als robust nachgewiesen¹⁸⁹. Die resultierende NNT-Kurve (Abb. 3.1) demonstriert, wie die NNTs mit sinkenden Ausgangsrisiken steigen und bei kleinen Risiken schließlich exponentiell zunehmen.
- **Umwandlung von Odds Ratio in NNT**. Für die Darstellung des Therapieeffekts aus der Meta-Analyse (ausgedrückt in OR) bei unterschiedlichen Ausgangsrisiken der Patienten wurde eine Umrechnung von Odds Ratio in NNT vorgenommen (Tab.3.4).

Tab. 3.4: Berechnung der NNT aus Ausgangsrisiko und Odds Ratio der Intervention

Anwendungsbeispiel: Behandlungsnutzen: OR_{H2-Blockers}: 0.58; Patientenrisiko_{GI-Blutung}: 2%

▪ Ausgangsrisiko (AR) = Ereignisrate der Kontrollgruppe	▪ AR _{GI-Blutung} = 2% = 0.02
▪ Ausgangs-Odds (AO) = $\frac{AR}{1 - AR}$	▪ AO = $\frac{0.02}{1 - 0.02} = 0.0204$
▪ Behandlungs-Odds (BO) = AO x Odds	▪ BO = 0.0204 x 0.58 = 0.0118
▪ Behandlungsrisiko (BR) = $\frac{BO}{1 + BO}$	▪ BR = $\frac{0.0118}{(1 + 0.0118)} = 0.0117$
▪ Risiko-Differenz (RD) = AR - BR	▪ RD = 0.02 - 0.0117 = 0.0083
▪ NNT = 1/RD	▪ NNT = $\frac{1}{0.0083} = 120$ Patienten

- Die **Präzision der Schätzung** für den Behandlungseffekt (ohne Berücksichtigung der Möglichkeit von Bias) erfolgt über das Konfidenzintervall (KI): In dem Beispiel der Ulkusprophylaxe mit einer OR von 0.58 reicht das Konfidenzintervall von 0.42 – 0.79¹⁶⁴.
- Die **Behandlungsschwelle** ausgedrückt als Schwellen-NNT beschreibt, bis zu welchen NNTs eine Behandlung als angemessen erscheint. Die Werte unterhalb der Schwelle liegen hier im Behandlungsbereich. In unser Modell sind neben der Wirksamkeit der Maßnahme weitere Determinanten eingeflossen, wie z.B. die Kosten von Zielereignis, Nebenwirkungen und Behandlung, die Häufigkeit des Auftretens von Nebenwirkungen (hepatische oder ZNS-Toxizität) und die Bedeutung, die Patienten und / oder Ärzte dem Auftreten eines Zielereignisses oder von Nebenwirkungen beimessen¹⁹⁰. Diese Determinanten werden in der Regel intuitiv verwendet. Dabei kam man für das Beispiel H₂-Blocker zur Prävention von GI-Blutungen auf eine NNT von 150.
- Die **Größenordnung von Bias** in randomisierten kontrollierten Studien durch unzureichenden Einsatz protektiver methodischer Verfahren (wie z.B. maskierte Randomisierung oder Verblindung der Endpunktbeurter) beläuft sich auf eine durchschnittliche Überschätzung des Behandlungseffekts um 40% bzw. 17%^{191; 192}. Verzerrungen dieser Größenordnung wurden für die Simulation herangezogen.

3.2.3 Ergebnisse

3.2.3.1 Individualisierung der Behandlung

H₂-Blocker sind hochwirksame Medikamente zur Prophylaxe von Stressulzera und GI-Blutungen. Abb. 3.1 gibt die NNT-Kurve mit einer Behandlungsschwelle von NNT (150) für eine Stressulcusprophylaxe bei unterschiedlichen Ausgangsrisiken der Patienten wieder. NNT-Kurve und Behandlungsschwelle schneiden sich bei einem Patientenrisiko von 1,6%. Nach diesem einfachen Modell ist somit für alle Patienten, deren Risiko für eine GI-Blutung ohne Behandlung höher als 1,6% ist, eine Prophylaxe mit H₂-Blockern indiziert. Diese Situation trifft für beide Patienten zu. Im Falle des Hochrisikopatienten A müssen nur 30 Patienten behandelt werden, um eine klinisch bedeutsame

GI-Blutung zu verhindern, während bei Niedrigrisikopatienten für den gleichen Effekt 120 Patienten behandelt werden müssen.

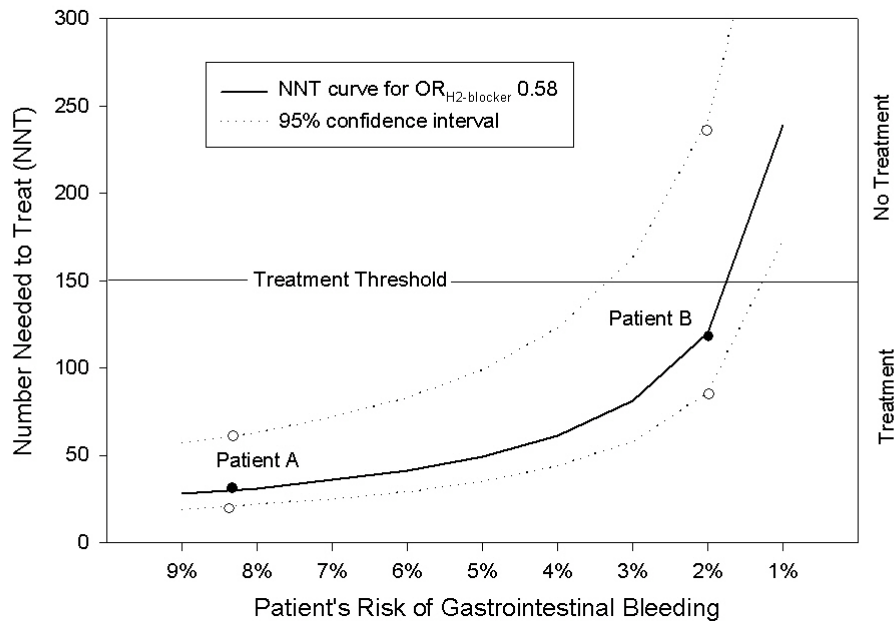


Abb. 3.1: NNT-Kurve für kritisch kranke Patienten mit unterschiedlichem Risiko für GI-Blutung, die mit H₂-Blocker behandelt werden (OR 0.58). Patient (A) ist ein Hochrisikopatient, Patient (B) ein Niedrigrisikopatient. Das Konfidenzintervall um die NNT-Kurve beschreibt die Genauigkeit des geschätzten Behandlungsnutzens. Die daraus resultierende Stärke der Behandlungsempfehlung ist für Hochrisikopatient A viel eindeutiger, da das Konfidenzintervall die Behandlungsschwelle nicht überschreitet, während sich das Konfidenzintervall bei Patient B über die Schwelle ausdehnt und somit Unsicherheit besteht, ob der Nettonachteil nicht den Nettonutzen überwiegt.

3.2.3.2 Fehlende Präzision bei der Schätzung des Behandlungseffekts

Der Schätzwert für die Wirksamkeit der Stressulkus-Prophylaxe (OR: 0.58) stammt aus einer Meta-Analyse mit 20 Placebo-kontrollierten Studien und ist mit einem 95%-Konfidenzintervall von 0.42 und 0.79, in dem sich der wahre Wert der Wirksamkeit mit 95% Wahrscheinlichkeit befindet, relativ präzise. Durch die Angabe eines Konfidenzintervalls um die berechneten NNTs¹⁹³ (Abb. 3.1) lässt sich die Schätzungsunsicherheit auch bei der Bewertung der H₂-Blocker-Gabe für die beiden Patienten ausdrücken: Für den Hochrisikopatienten (A) erstreckt sich das 95% KI_{NNT} von 21 Patienten bis 61 Patienten, d.h. bei einem Ausgangsblutungsrisiko von 8,4% müssen im besten Fall nur 21 Patienten, um ungünstigsten Fall 61 Patienten behandelt werden, um eine klinisch relevante GI-Blutung zu

verhindern. Für den Niedrigrisikopatienten (B) spannt sich das 95% KI_{NNT} von 87 Patienten bis 242 Patienten, d.h. im ungünstigsten Falle müssen 242 Patienten behandelt werden, um eine GI-Blutung zu verhindern. Dieses Intervall überschreitet die festgelegte Behandlungsschwelle von 150 Patienten und demonstriert damit die Unsicherheit, ob der Patient von der Behandlung noch wirklich profitiert. Entsprechend fällt die Empfehlung, Patienten B zu behandeln weniger stark aus als die Empfehlung für Patienten A. Die Unsicherheit, die über den wahren Behandlungseffekt besteht, muss daher bei Behandlungsempfehlungen wie Leitlinien zum Beispiel in Form von Empfehlungsstärken („Grades of Recommendations“¹⁵⁹) zum Ausdruck gebracht werden.

3.2.3.3 Die Auswirkung von Bias in Forschungsergebnissen auf ärztliche Entscheidungen

Werden Studien nicht mit ausreichender Sorgfalt entworfen und durchgeführt, können bei der Schätzung des wahren Behandlungseffekts Verzerrungen auftreten, durch die der Behandlungseffekt systematisch über- oder unterschätzt wird. Die zugrundeliegende Meta-Analyse hatte in der Qualitätsbewertung der Primärstudien festgestellt, dass die eingeschlossenen Studien weniger als die Hälfte der maximal möglichen Qualitätspunkte erreichen (im Schnitt 5,2/12 Punkten). Methodisch mittelmäßig bis minderwertige Studien sind besonders anfällig für Bias. Im Folgenden soll beleuchtet werden, inwieweit sich verschiedene Konstellationen von Bias und das Zusammentreffen bestimmter klinischer Situationen auf ärztliche Entscheidungen auswirken können.

3.2.3.3.1 Ausgeprägte Bias bei großem Behandlungseffekt

Erstes Szenario: Verzerrung eines real großen Behandlungseffekts durch ausgeprägte Bias im Studiendesign

Eine Überschätzung des Behandlungseffekts um durchschnittlich 40% kann z.B. durch eine unmaskierte Randomzuteilung der Studienteilnehmer auftreten¹⁹¹. Die bereits hochwirksame Behandlung mit H₂-Blockern (hier: OR_{H₂-Blockern}: 0.58) erscheint irrtümlicherweise noch effektiver, nämlich als OR_{H₂-Blockern+40%Bias} von 0.35 und kommt damit außerhalb des KI_{OR-0.58} von 0.42 – 0.79 zu

liegen. Die zugehörige $NNT_{40\%-\text{Bias}}$ -Kurve verläuft unterhalb der unverzerrten $NNT_{OR-0.58}$ -Kurve (Abb.3.2) und kreuzt erst bei einem Ausgangsrisiko von 1% die Behandlungsschwelle. Damit wird die Prophylaxe mit H_2 -Blockern auch für Patienten mit einem Ausgangsrisiko zwischen 2% und 1% als angemessen betrachtet. Dieser Unterschied erscheint im ersten Moment banal. Da aber wesentlich mehr Patienten ein niedriges Risiko aufweisen, würde die erweiterte Indikationsstellung auf eine erheblich größere Patientenzahl zutreffen.

In unserem Szenario sinkt die NNT für den Hochrisikopatienten auf 19 (statt bisher 30) und für den Niedrigrisikopatienten auf 77 (statt bisher 120). Trotz erheblicher Überschätzung der Wirksamkeit ändert sich bei dem großen Behandlungseffekt für keine der beiden Risikokonstellationen die Indikation zur Behandlung.

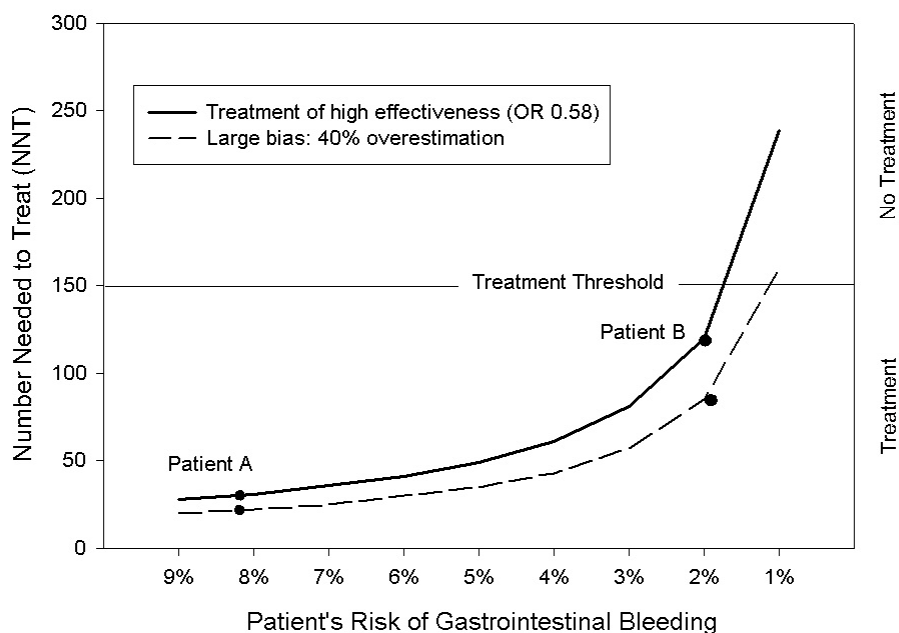


Abb.3.2: Die Auswirkung von Bias auf die klinische Entscheidung bei großem Behandlungseffekt.

Bei großem Behandlungsnutzen ist es selbst bei Niedrigrisikopatienten wenig wahrscheinlich, dass ein ausgeprägter Biaseffekt (40% Überschätzung des wahren Behandlungsnutzens) die ärztliche Entscheidung beeinflusst.

3.2.3.3.2 Ausgeprägte (bzw. mäßig große) Bias bei moderatem Behandlungseffekt

Zweites Szenario: Verzerrung eines moderaten Behandlungseffekts durch ausgeprägte (bzw. mäßig große) Bias im Studiendesign

Beim zweiten Szenario wird angenommen, dass die Prophylaxe mit H₂-Blockern weniger wirksam sei und bei unverzerrter Schätzung nur eine Reduktion der GI-Blutungen auf eine OR von 0.8 erzielt werden könnte (Abb.3.3). Eine Vielzahl der zur Verfügung stehenden medizinischen Behandlungsformen haben eine Wirksamkeit in dieser Größenordnung. In dieser Situation schneiden sich NNT_{OR=0,8}-Kurve und Behandlungsschwelle bei einem Ausgangsrisiko von 3,5% und man würde die Behandlung nur Patienten mit einem Risiko > 3,5% empfehlen. Des weiteren wird angenommen, dass in diesem Szenario die Wirksamkeit der Prophylaxe durch Bias einmal um 40% (wie in dem ersten Szenario) und einmal nur um 20% überschätzt wird. Beide Verzerrungen haben für den Hochrisikopatienten A keine Konsequenzen: In allen drei Konstellationen (ohne Verzerrung, mit moderater und mit großer Verzerrung) befinden sich die zugehörigen NNTs im Behandlungsbereich in großem Abstand zur Schwelle. Der Niedrigrisikopatient B dagegen liegt bei unverzerrter Schätzung mit einer NNT von 254 deutlich oberhalb der Behandlungsschwelle im nicht behandlungswürdigen Bereich, da der zu erzielende Nutzen im Verhältnis zu Nebenwirkungen und Kosten für ihn als zu gering erachtet wird. Selbst bei einer relativ gering ausgeprägten Bias in den Studien von nur 20% (was bei unverblindeten Studien leicht vorkommen kann¹⁹¹) würde der scheinbare Nutzen der Behandlung mit einer NNT von 141 auf die andere Seite der Schwelle rutschen und der Patient irrtümlich die Prophylaxe erhalten. Auch hier werden von der Fehlentscheidung wieder eine große Gruppe von Patienten getroffen.

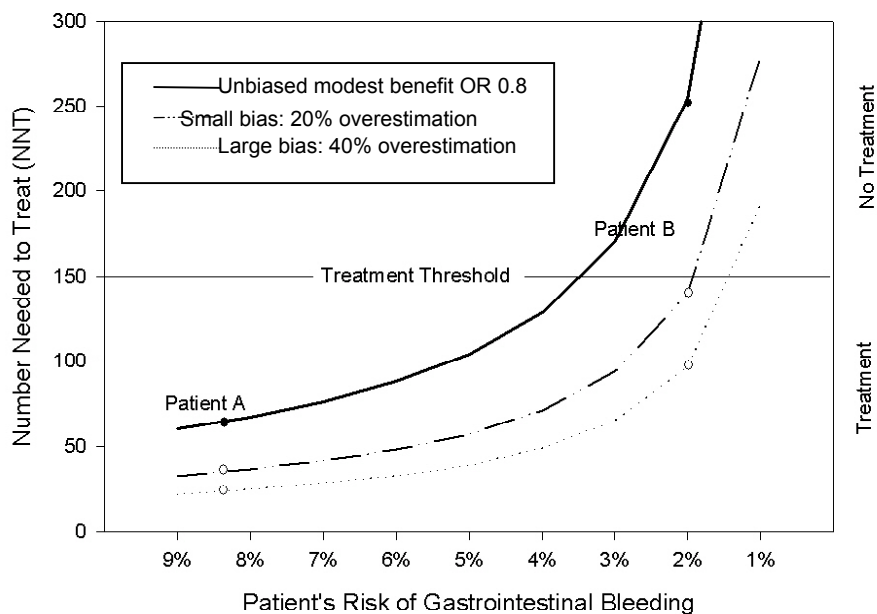


Abb. 3.3: Die Auswirkung von Bias auf die ärztliche Entscheidung bei mäßig großem Behandlungseffekt

Während die ärztliche Entscheidung bei Hochrisikopatienten (A) selbst bei ausgeprägter Bias nicht beeinflusst wird (40% Überschätzung des Behandlungsnutzens), können bei Niedrigrisikopatienten und geringem Behandlungsnutzen bereits kleine Biaseffekte (20% Überschätzung des Behandlungsnutzens) zu ärztlichen Fehlentscheidungen führen.

3.2.4 Schlussfolgerungen

3.2.4.1 Konsequenzen für den behandelnden Arzt

Mit der vorliegenden Simulation konnten wir zeigen, dass sich bei gleichem Ausmaß an Bias die NNT-Kurve im Niedrigrisikobereich und bei mäßig effektiven Maßnahmen wesentlich stärker verändert als im Hochrisikobereich und bei hochwirksamen Maßnahmen. In der ersten Konstellation kommen die NNTs in der Regel nahe an einer impliziten oder expliziten Behandlungsschwelle zu liegen und selbst geringe Biaseffekte können die NNTs aus dem „Nicht“-Behandlungsbereich über die Schwelle in den Behandlungsbereich heben. Da die meisten Interventionen nur kleine bis moderate Effekte (20-25% RRR) aufweisen und es wesentlich mehr Patienten mit niedrigerem Risiko gibt, ist die Anzahl an Patienten, die von Fehlentscheidungen auf der Grundlage methodisch unsauber durchgeführter Studien getroffen werden, in der Regel gross (Tab. 3.5). Praktizierende Ärzte müssen

deshalb bei Patienten eine wesentlich größere Sensibilität für potentiell verzerrte Studienergebnisse haben, wenn die Therapieeffekte gering sind.

Tab. 3.5: Potential für medizinische Fehlentscheidungen durch Bias

Empfehlung bei ...	Mäßig effektive Behandlung		Hochwirksame Behandlung	
	Hohes Risiko (Patient A)	Niedriges Risiko (Patient B)	Hohes Risiko (Patient A)	Niedriges Risiko (Patient B)
... Therapieeffekten ohne Bias	Behandlung	Keine Behandlung	Behandlung	Behandlung
... großem Biaseffekt	Behandlung	Behandlung: Fehlerhafte Entscheidung	Behandlung	Behandlung
... mäßig großem Biaseffekt	Behandlung	Behandlung: Fehlerhafte Entscheidung	Behandlung	Behandlung

3.2.4.2 Bias aus Beobachtungsstudien

Die aufgeführten Szenarien beruhen ausschließlich auf den empirischen Kenntnissen über die Auswirkungen von Bias in randomisierten Studien. In Beobachtungsstudien können durch die fehlende Randomisierung noch wesentlich größere Behandlungseffekte vorgetäuscht oder Effekte gefunden werden, die gar nicht existieren (s. Diskussion 2.2). Ein jüngstes Beispiel ist der Einfluss der kombinierten Hormonersatztherapie auf das kardiale Risiko von postmenopausalen Frauen. In Kohortenstudien war für Östrogene wiederholt ein protektiver Effekt auf das KHK-Risiko postmenopausaler Frauen gefunden worden¹⁹⁴. Weder bei Frauen mit vorbestehender KHK¹⁵⁵ noch bei gesunden postmenopausalen Frauen konnte dieser Effekt in randomisierten Studien bestätigt werden²⁵. Sehr ernst zunehmen ist, dass die Behandlung bei den Frauen sogar zu ernsthaftem Schaden führte. Diese Unsicherheit um die Zuverlässigkeit der Aussagen und damit verbunden das erhebliche Potential für ärztliche Fehlentscheidungen begründen und bestärken die niedrigere Klassifizierung von Beobachtungsstudien in den Evidenzstufen von Interventionsstudien^{159; 195}.

3.2.4.3 Implikationen für die Praxis

Nur ein solides Verständnis der Wechselwirkungen von Risiko, Behandlungsnutzen und Bias gewährleistet eine differenziertere „angemessene“ Patientenversorgung, wie sie derzeit überall gefordert wird. Dennoch werden diese Zusammenhänge in der Patientenversorgung nicht ausreichend wahrgenommen und umgesetzt. Eine praktische Möglichkeit der raschen Implementierung sind Leitlinien, die von Ärzten zunehmend als Unterstützung ihrer täglichen Arbeit verstanden werden. Allerdings sind Leitlinien wegen der oftmals beobachteten Beliebigkeit der Erstellung und der z.T. nicht zu übersehenden Instrumentalisierung für eigene Interessen in Misskredit geraten (z.B. die > 2000 expertenbasierten AWMF-Leitlinien, www.awmf-online.de). Transparenz in der Erstellung und Glaubwürdigkeit bei den Empfehlungen, entsprechend dem Anspruch evidenzbasierter Leitlinien, sind Voraussetzung für eine allgemeine Akzeptanz. Dabei kann ein Clearingverfahren diesen Prozess unterstützen.

Im folgenden Abschnitt wird beschrieben, wie mit einem Clearingverfahren für Hypertonieleitlinien die deutschen mit den internationalen Hypertonieleitlinien anhand von EbM-Kriterien standardisiert verglichen („Benchmarking“), Stärken und Problembereiche identifiziert und Empfehlungen zu Eckpunkten einer nationalen Hochdruckleitlinie abgegeben wurden.

3.3 Der Leitlinien-Clearingbericht Hypertonie (Eigene Ergebnisse)

3.3.1 Das Leitlinien – Clearingverfahren der Ärztlichen Zentralstelle Qualitätssicherung

Um die Transparenz des unübersichtlich gewordenen Angebotes an Leitlinien in Deutschland zu verbessern, wurde 1999 von den Spitzenverbänden des Gesundheitswesens an der Ärztlichen Zentralstelle Qualitätssicherung (ÄZQ) in Köln ein „LL-Clearingverfahren“ eingerichtet, das unter anderem vorhandene Leitlinien nach expliziten Kriterien und einer standardisierten Vorgehensweise auf ihre Transparenz, Wissenschaftlichkeit, Praktikabilität und Wirtschaftlichkeit bewerten und gegebenenfalls Empfehlungen zur Verbesserung aussprechen sollte¹⁹⁶. In den Vergleich sollten nationale und internationale Leitlinien einbezogen und an den internationalen Vorstellungen über

Leitlinienqualität^{197; 198}, die sich stark an den Konzepten der evidenzbasierten Medizin orientieren, gemessen werden.

Die Hypertonie gehört zu den prioritären Gesundheitsproblemen¹⁹⁹, für die ein Bedarf an stärkerer Standardisierung der Versorgung durch Leitlinien gesehen wird. Die Erarbeitung des Clearingberichts Hypertonie war das erste von der ÄZQ durchgeführte LL-Clearingverfahren mit einer sechsköpfigen Expertengruppe aus Leitlinienanwendern der betroffenen Fachdisziplinen im ambulanten / stationären Versorgungsbereich und Methodikern. In diesem Gremium habe ich als Vertreterin der Fächer Klinische Epidemiologie/ Innere Medizin mitgewirkt.

3.3.2 Zielsetzung des Leitlinien-Clearingverfahrens Hypertonie

Die Ziele des Leitlinienverfahrens waren eine

1. umfassende Recherche nach nationalen und internationalen Hypertonieleitlinien
2. kritische Bewertung nach EbM-Kriterien
3. Erstellung eines Clearingberichts, ggf. mit Formulierung von Empfehlungen für eine nationale evidenzbasierte Hypertonieleitlinie

3.3.2.1 Methode

Das Clearingverfahren umfasst drei Schritte:

- Umfassende Identifizierung der verfügbaren deutsch- und englischsprachigen LL in vier allgemeinen Literaturdatenbanken (Medline, HealthStar, Embase, SciSearch) sowie 25 fachspezifischen und 33 fachübergreifenden LL-Datenbanken
- Formale Qualitätsbewertung der Bereiche „LL-Entwicklung“, „Inhalt/Form“ und „Anwendung/Implementierung“ mit einem nach internationalen Standards¹⁹⁸ entwickelten Instrument (s. Anhang 3.1) und Selektion der hochwertigsten LL
- Inhaltliche Bewertung und Erstellung des Clearingberichts mit strukturellen und inhaltlichen Eckpunkten für eine nationale LL Hypertonie durch den Expertenkreis (weitere methodische Details im Clearingbericht¹¹⁶)

3.3.2.2 Ergebnisse

3.3.2.2.1 Formale Bewertungen

Mit einer strukturierten Literatursuche wurden 548 Publikationen gefunden, die auf 132 Zitate mit Abstract eingeschränkt wurden. Elf von 35 LL (mit Basiskriterien Hypertonie allgemein, deutsche oder englische Sprache, überregionale LL, s. Anhang 3.2) erfüllten auch die endgültigen Einschlusskriterien („Entwicklung und Inhalt genauso gut wie oder besser als die Hypertonie-LL der AWMF“), darunter die beiden deutschsprachigen Leitlinien^{200; 201}. Die Ergebnisse der formalen Bewertung mit den drei Teilbereichen „Entwicklung“, „Inhalt“, „Anwendung/Umsetzung“ sind in Tabelle 3.6 zusammengefasst.

Tab. 3.6: Formale Bewertung der elf eingeschlossenen Leitlinien

Leitlinie	Literaturzitate	Entwicklung (max: 17 Pkt)	Inhalt (max: 17 Pkt)	Anwendung (max: 6 Pkt)	Gesamt (max: 40 Pkt)
CDN 99 ²⁰²	145	15 (EB)	16	4 (TR)	35
CDN 99 ²⁰³	463	14 (EB)	17	3	34
JNC 97 ²⁰⁴	254	14 (EB)	15	5	34
ICSI 99 ²⁰⁵	59	11 (EB)	16	6	33
AkdÄ* 98 ²⁰¹	33	11 (EB)	15	4	30
BHS 99 ²⁰⁶	170	8 (EB)	14	6	28
HSSA 95 ²⁰⁷	55	7	16	0	23
DVA 96 ²⁰⁸	25	8	14	0	22
WHO 99 ²⁰⁹	199	6	14	0	21
NZ 97 ²¹⁰	24	6	14	0	20
AWMF* ²⁰⁰	15	5	7	4 (TR)	16

EB: Verknüpfung: Empfehlung / klassifizierte Evidenz

TR: LL ist Grundlage für Trainingsmaßnahmen

* Deutsche Leitlinien

Auch wenn man die absoluten Werte der Scores mit Vorsicht interpretieren sollte, zeigt sich, dass bei „Inhalt“ zwischen den Leitlinien große Ähnlichkeiten vorlagen, während vor allem bei „Entwicklung“ und „Anwendung/Implementierung“ z.T. erhebliche Unterschiede zu Tage traten. Die sechs Leitlinien mit dem höchsten Gesamtscore erreichten ihre hohe Bewertung vor allem durch Begründung der

einzelnen Empfehlungen mit klassifizierter Evidenz und durch Aussagen zu den Bereichen „Implementierung/Anwendung“.

3.3.2.2 Inhaltliche Empfehlungen für eine nationale Leitlinie Hypertonie

Da keine der LL vollständig den Kriterien einer praktikablen evidenzbasierten LL entsprach, formulierten wir strukturelle und inhaltliche Eckpunkte für eine nationale LL Hypertonie, die das Krankheitsbild Hypertonie aus dem Gesichtspunkt der Versorgung umfassend abdecken sollte. Tab 3.7 fasst die Eckpunkte zusammen.

Tab. 3.7 : Inhaltliche Schwerpunkte einer nationalen Leitlinie Hypertonie

1. Einführung: Definition Ziele, Adressaten, Epidemiologie, Versorgungsproblematik,	9. Nicht medikamentöse Therapie
2. Blutdruckmessung	10. Pharmakotherapie
3. Anamnese und Untersuchung	11. Follow-up, Schulung, Motivation, Compliance
4. Case-finding / Screening	12. Komorbidität, spezifische Patientengruppen
5. Schnittstellen in der Versorgung	13. Primärprävention
6. Risikoeinschätzung	14. Qualitätssicherung / -management
7. Weiterführende Diagnostik	15. Disseminierung und Implementierung
8. Therapieindikationen und –ziele	16. Offene Fragen / Forschungsthemen

Durch Verknüpfung der Kernaussagen mit klassifizierter Evidenz sollte transparent gemacht werden, mit welcher Art und Qualität von wissenschaftlichen Studien die Empfehlungen unterstützt wurden. Ein weiteres wichtiges Anliegen der Empfehlungen des Clearingberichts war eine stärkere Betonung des Zusammenhangs von kardiovaskulärem Gesamtrisiko und Wirksamkeit und die Folgen für die Behandlungsindikation beim individuellen Patienten. Zusätzlich sollte beschrieben und belegt werden, welche Patientengruppen wann und in welchem Ausmaß von einer Behandlung profitieren. Vorbild für die zuletzt genannten Empfehlungen waren die Risikotabellen der British Hypertension Society²⁰⁶

oder die New Zealand Guidelines²¹⁰, mit denen sich das individuelle Patientenrisiko unter simultaner Berücksichtigung multipler Risikofaktoren berechnen und der individuelle Behandlungsnutzen darstellen ließ. Die Empfehlungen, die vor allem bei den Eckpunkten ‚Risikoeinschätzung‘, Therapieindikation / -ziele sowie „Qualitätssicherung“ ausgesprochenen wurden, sind in Tabelle 3.8 zusammengefasst.

Tab. 3.8: Zusammenfassung der Empfehlungen des Clearingberichts Hypertonie¹¹⁶ zu Risikoeinschätzung und Präsentation der Wirksamkeit in einer nationalen Leitlinie Hypertonie

1. Betonung einer strukturierten Ermittlung des Risikoprofils wegen der Unzuverlässigkeit intuitiver Schätzungen.
2. Quantitative Abschätzung des Gesamtrisikos (für die zerebrale, kardiale, renale und vaskuläre Morbidität und Mortalität) auf der Grundlage der bekannten Risikofaktoren mit Hilfsmitteln (wie Risikotafeln oder computergestützte Risikoberechnung), die im Rahmen der LL zur Verfügung gestellt werden sollen
3. Darstellung des Risikos für Hypertonie-Folgeerkrankungen und die zu erwartende Risikosenkung (Behandlungsnutzen) in verständlicher Form (ARR und NNTs/ Jahr)
4. Partizipation des Patienten an der Entscheidungsfindung durch Darstellung des zu erwartenden Behandlungsnutzens in einer für den Patienten verständlichen Form
5. Entwicklung von Instrumenten zur Kommunikation von Risiko und Behandlungsnutzen an den Patienten
6. Vermittlung der Grundbegriffe zur Darstellung von Risiko und Nutzen (RRR, ARR, NNT) sowie die Bestimmung des kardiovaskulären Risikos im Rahmen der Qualitätssicherung

NNT: Number Needed to Treat; ARR: Absolute Risikoreduktion; RRR: Relative Risikoreduktion

Angesichts der zunehmenden Orientierung praktizierender Ärzte an den Leitlinien der Fachgesellschaften würde eine evidenzbasierte Leitlinie (z.B. nach dem Vorbild der schottischen SIGN-Leitlinien www.sign.ac.uk) die Gelegenheit bieten, Ärzte anhand des praktischen Beispiels der Versorgung von Hypertonikern an die Grundkonzepte der evidenzbasierten Medizin heranzuführen. Dabei ist die klinisch-epidemiologische Evidenz im Bereich Hypertonie besonders gut fundiert und steht vielfach nach EbM-Kriterien aufgearbeitet bereits zur Verfügung.

3.3.3 Schlussfolgerungen

Verschiedene experimentelle Studien sowie unsere Simulationsstudie belegen das Potential für bedeutsame Fehlentscheidungen in der Versorgung, deren Konsequenzen zumindest einen Beitrag für die im Sachverständigengutachten beschriebene Über-, Unter- und Fehlversorgung liefern können. Um diesen Fehlentwicklungen angemessen zu begegnen, benötigen Ärzte ein besseres Verständnis für quantitative (klinisch-epidemiologische) Informationen, wie z.B. den Zusammenhang von Risiko, Wirksamkeit und Nutzen. Für eine rasche flächendeckende Verbreitung sind evidenzbasierte Leitlinien ein erstes Transportmittel dieser Kenntnisse. Ohne ein Verständnis für die dahinterstehenden Konzepte und bei dem derzeitigen offensichtlichen Reglementierungsdruck besteht eine nicht unerhebliche Gefahr, dass das Potential der Leitlinien als echtes Hilfsmittel für Entscheidungen von den Ärzten nicht wahrgenommen wird. Ein ganzes System zum Umdenken zu motivieren ist ein schwieriger, langsamer Prozess, bei dem viele Schrauben gleichzeitig bewegt werden müssen. Eine stärkere Dissemination der EbM-Konzepte, z.B. im Rahmen von EbM-Kursen (Kapitel 4) oder durch konkrete Praxisprojekte (Kapitel 5) sind dabei essentielle flankierende Maßnahmen.

4 Evidenzbasierte Medizin – Neue Fertigkeiten?

Aufbau des Kapitels

Kurse zur Vermittlung der Methoden und Inhalte der evidenzbasierten Medizin werden inzwischen weltweit angeboten. In diesem Kapitel geht es um die Frage, ob durch die systematische Vermittlung von EbM z.B. im Rahmen von EbM-Kursen nachweislich neue Kompetenzen erworben werden.

Die Einführung (4.1) beschreibt, wie die strukturierte Vermittlung und Umsetzung von EbM die modernen Vorstellungen über das Erwachsenenlernen und das lebenslange Lernen erfüllen, was ihre Attraktivität für die ärztliche Fort- und Weiterbildung erhöht, wie aus diesen Gründen die EbM in die ärztlichen Weiterbildungsprogramme angelsächsischer Länder integriert wurde und welche Qualifikationsmöglichkeiten für EbM sich im Vergleich dazu in Deutschland bieten. Die wiederholt vorgebrachte Kritik über den fehlenden Nachweis des Nutzens von EbM-Kursen war für uns der Anlass, dieser Frage systematisch nachzugehen (eigene Arbeiten: 4.2): Dabei definierten wir zuerst die Inhalte der Berliner EbM-Kurse in einem Gegenstandskatalog, dem „Berliner Gegenstandskatalog“ (4.2.1) und entwickelten und validierten anschliessend ein Instrument zur Messung des Zuwachses an Wissen und Fertigkeiten (4.2.2). Mit diesem Instrument konnten wir in einer dreijährigen Studie über den Nutzen der EbM-Ausbildung einen klinisch bedeutsamen Zugewinn an Wissen nachweisen (4.2.3).

Eigene eingeschlossene Publikationen zum Thema

- Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer H-H, Kunz R (senior author). Do short courses in evidence-based medicine improve knowledge and skills? Validation of the Berlin questionnaire and before and after study of EBM training. *BMJ* (accepted: 21.10.2002)
- Kunz R, Fritsche L, Neumayer H-H. Quality assurance criteria for continuing education in evidence based medicine. *ZaeFQ* 2001; 95: 371-375
- Kunz R, Langenberg C, Fritsche L. Fortbildungsangebote in EbM. Fragen und offene Wünsche. *ZaeFQ* 2001; 95: 259
- Kunz R, Fritsche L. Evidenzbasierte Medizin. Von der Evidenz-Produktion zu ihrer Nutzung. *Z Rheumatol* 2000; 59: 38-40
- Kunz R. Evidence-based Medicine im klinischen Alltag. *Journal für Anästhesie und Intensivbehandlung* 2000; I: 105-106
- Fritsche L, Neumayer H.-H., Kunz R. Vom Konzept in die Praxis. Umsetzbarkeit eines strukturierten Programms für einen Grundkurs in Evidence-based Medicine. *ZaeFQ* 1999; 93: 73-78

- Kunz R, Weingart O, Jonitz G, Engelbrecht J, Antes G, Ollenschläger G. Evidenzbasierte Medizin. Grundlage ärztlichen Handelns. Deutsches Ärzteblatt 2002; 99: A2685-A2686
- Kunz R, Fritsche L, Budde K. Unter der Lupe von EBM. In: Perleth M, Antes G (Hrsg.). Evidenzbasierte Medizin. Wissenschaft im Praxisalltag. MMV München 2003 (3. Aufl.); 27-37
- Khan KS, Kunz R, Kleijnen J, Antes G. Systematic Reviews to Support Evidence-based Medicine: How to review and apply findings of health care literature. London: The Royal Society of Medicine Press, 2002 (In press)
- Guyatt G, Rennie D. (Ed.) Users' Guides to the Medical Literature. JAMA and Archives; American Medical Association. 2001.
- Kunz R, Ollenschläger G, Raspe H-H, Jonitz G, Kolkmann FW (Hrsg.). Lehrbuch der Evidenzbasierten Medizin. EBM in Klinik und Praxis. Dt. Ärzteverlag 10/2000
- Sackett DL, Richardson WS, Rosenberg W, Haynes RB (Deutsche Herausgeber: Kunz R, Fritsche L). Evidenzbasierte Medizin. München, Bern, Wien, New York. 1999

4.1 EbM – ein Modell für lebenslanges Lernen in der Medizin

Mit den neuen gesetzlichen Vorgaben läuft die Einführung von EbM in Deutschland nahezu ausschließlich über die Erstellung von sog. „evidenzbasierten“ Vorgaben, als Leitlinien oder im Rahmen von Disease-Management-Programmen. Damit erhofft sich der Gesetzgeber u.a. eine Verbesserung der Patientenversorgung. Systematische Untersuchungen zur Qualität der Patientenversorgung haben für Deutschland wichtige Bereiche mit Über-, Unter- und Fehlversorgung der Bevölkerung aufgezeigt, wie der kürzlich vorgelegte Bericht des Sachverständigenrats ausführlich belegt¹⁸⁶.

Für diese Defizite wurden unter anderem Fehlentwicklungen in der Fort- und Weiterbildung der Ärzte verantwortlich gemacht¹⁸⁶. Es ist dokumentiert, dass das Wissen der Ärzte selbst für häufige Krankheitsbilder²¹¹ nach Abschluss der Ausbildung kontinuierlich zurückgeht²¹²⁻²¹⁴, ohne dass der Kompetenzverlust erkannt oder auf erkannten Informationsbedarf angemessen reagiert wird²¹⁵⁻²¹⁷. Die traditionelle Form der Fortbildung durch Publikationen und Vorträge scheint für eine erfolgreiche zeitnahe Integration von neuen Forschungsergebnissen in die Versorgung unzureichend²¹⁸⁻²²⁰. Entsprechend wird auch die Gefahr ärztlicher Inkompetenz inzwischen öffentlich diskutiert^{213; 214}.

Obwohl die Anforderungen an die Aus- und Weiterbildungsangebote zunehmend steigen²²¹, vermitteln sie den Ärzten bisher nicht in ausreichendem Maße die Fertigkeiten für ein lebenslanges selbstbestimmtes Lernen, die sie benötigen, um auf die sich rasch ändernden Inhalte und

Anforderungen reagieren zu können. EbM-Kurse sind eine Möglichkeit, die dazu notwendigen Methoden und Fertigkeiten zu erlernen.

4.1.1 EbM-Kurse nach dem McMaster-Format (Didaktik der EbM)

Die evidenzbasierte Medizin ist in Inhalt und Form ein Produkt der McMaster-University: Problemorientiertes Lernen (POL) und klinisch-epidemiologische Denkweise werden zusammengeführt, um dem Arzt bei der Lösung spezifischer Patientenprobleme zu helfen²²². Die an der McMaster University entwickelten „Evidence-based Medicine“-Workshops sind der Prototyp für die Vermittlung von EbM. Sie wurden von D.L. Sackett in Oxford weitergeführt und finden inzwischen weltweit statt. Im Folgenden sollen die zugrundeliegenden didaktischen Prinzipien kurz skizziert werden, eine detaillierte Beschreibung findet sich in den einschlägigen Publikationen²²²⁻²²⁴.

Der Inhalt: Typischerweise findet das Lernen in Kleingruppen von fünf bis sieben Personen statt, die von ein bis zwei Tutoren betreut werden. Anhand eines typischen Patientenproblems entwickeln sie gemeinsam die fünf Schritte der „EbM“-Problemlösung: 1) Umwandlung des Problems in eine fokussierte Frage, 2) Literatursuche in gängigen Datenbanken und Identifizierung einer geeigneten Studie, 3) kritische Bewertung der Studie nach klinisch-epidemiologischen Kriterien und Beurteilung der klinischen Relevanz der Ergebnisse, 4) Anwendung der Ergebnisse auf den Patienten in Abwägung des patientenspezifischen Risikos, der Behandlungsalternativen, Nebenwirkungen und Patientenvorstellungen und 5) selbstkritische Bewertung der eigenen Fertigkeiten, entsprechend dem Konzept des problemorientierten Lernens (POL). Nach diesem Schema werden Problemlösungen zu den großen Themen der ärztlichen Praxis erarbeitet: Therapie und Prävention, Diagnostik und Screening, Prognose und Nebenwirkungen, in Fortgeschrittenenkursen auch Themen wie Leitlinien, gesundheitsökonomische Analysen oder Entscheidungsanalysen. Da bei der klassischen EbM die Patientenversorgung im Mittelpunkt steht, werden die Kurse von Ärzten mit klinisch-epidemiologischer Ausbildung oder fundierten EbM-Kenntnissen betreut, die die Umsetzung in der Praxis aus eigener Erfahrung glaubwürdig und konstruktiv diskutieren können. Damit erhält die kritische Literaturbewertung auch ihren Stellenwert: nicht das Aufspüren der kleinsten methodischen Schwäche der Studie steht im Vordergrund, sondern eine qualifizierte Bewertung darüber, welchen

Beitrag diese Evidenz für die Entscheidungsfindung liefern kann. Die kritische Bewertung von Studien ist also ein wichtiger Bestandteil, aber kein Synonym von EbM.

Das Format: Entsprechend der Verwurzelung des Kurses im POL übernehmen die Teilnehmer wesentliche Verantwortung für Inhalt und Gestaltung und lernen primär mit dem „Wissen der Gruppe“. Die Tutoren, obwohl inhaltliche Experten, fungieren im Wesentlichen als Moderatoren und springen mit ihrem Wissen erst als letzte Instanz ein. Damit erfahren die Teilnehmer die Grundbegriffe des problemorientierten Lernens als festen Bestandteil einer evidenzbasierten Praxis. Das läuft nicht immer ohne Frustration ab, da nicht für jede Frage sofort eine Lösung parat steht, der Ablauf spiegelt aber die täglichen Erfahrungen eines jeden Teilnehmers wider. Der bewusste Umgang mit Wissenslücken ist eine wichtige Voraussetzung für die kompetente Integration von EbM in die tägliche Praxis²²².

Das Kursformat fordert von den Tutoren hohe inhaltliche und didaktische Kompetenzen, da sie spontan auf schwierige Situationen und Fragen reagieren müssen. Die Nähe zu den heute überall geforderten Fertigkeiten des lebenslangen Lernens^{222; 225; 225} ist dabei nicht zu übersehen.

Inzwischen gibt es viele Varianten des 5-Tage-McMaster-Kurses: Programme für akademische Halbtage²²⁶, 2-Tages-Kurse, 3½-Tages-Kurse wie den Berliner EbM-Kurs²²⁷ oder wöchentliche Kurse über einen längeren Zeitraum. Die wiederholt angeforderte systematische Evaluation der Kurse und Programme beginnt erst jetzt mit deutlicher Verzögerung^{228; 228-230}.

4.1.2 EbM in der Weiterbildung - Ausländische Erfahrungen

4.1.2.1 „EbM-Practitioners“: Modell der ärztlichen Weiterbildung an der McMaster-University

Die Umsetzbarkeit einer konsequenten evidenzbasierten Medizin mit ihren nicht unerheblichen Ansprüchen in die Routineversorgung ist immer wieder in Frage gestellt worden²³¹. Um die Machbarkeit zu untersuchen, wurde Anfang der 90er Jahre an der McMaster University das „Internal Medicine Residency Program“²³² in ein Ausbildungsmodell für „ebm-practitioners“ umstrukturiert, ein

mutiges Unterfangen unter den damaligen Rahmenbedingungen (die Cochrane Collaboration noch nicht gegründet, Computer wenig verbreitet, PubMed und Online Artikel nicht verfügbar). In einem speziellen „Faculty Development Program“ wurden neue Lernziele für die Residents formuliert, die gesamte Faculty in „Critical appraisal“ geschult und die Stationen flächendeckend mit Computer ausgestattet. Mit einem „EbM-Peer-Programm“ erprobte und reflektierte die Fakultät die eigenen Ausbildungsstrategien in der Routineversorgung. Wesentliche Voraussetzung für die erfolgreiche Umsetzung waren aber auch die solide finanzielle Ausstattung sowie die positive Einstellung und aktive Unterstützung, mit der die Fakultätsmitglieder das Modellprojekt mitgetragen haben.

Seitdem hat sich die Praxis der EbM erheblich weiterentwickelt. Mit der zunehmenden Verfügbarkeit vorbewerteter EbM-Ressourcen (wie die Cochrane Library²³³, Best Evidence²³⁴ oder Clinical Evidence²³⁵) ist die Bedeutung der selbstständigen Bewertung von Originalstudien etwas in den Hintergrund getreten, die anderen EbM-Fertigkeiten wie Identifizierung von Wissenslücken, Formulierung gezielter Fragen, Auffinden der vorbewerteten Evidenz und differenzierte Anwendung auf den Patienten haben an Bedeutung gewonnen. Mit dieser einfacheren Form können sich auch Ärzte mit weniger Interesse an vertieften Fertigkeiten der kritischen Studienbewertung zu kompetenten EbM-Nutzern weiterentwickeln²³⁶. Das bestätigte eine Umfrage bei Allgemeinärzten in Großbritannien nach EbM-Weiterbildungen, von denen 70% angaben, vorbewertete Primärstudien oder evidenzbasierte Praxisleitlinien zu verwenden, aber der selbstständigen Studienbewertung eher skeptisch gegenüberstanden²³⁷.

4.1.2.2 EbM in US-amerikanischen Residency-Programmen

EbM-Konzepte oder einzelne EbM-Elemente sind inzwischen fest in vielen US-amerikanischen Ausbildungsprogrammen integriert²³⁸. Mehr als 30% der Programme lehrten kritische Studienbewertung (78%), systematische Literatursuche (53%), das Formulieren fokussierter Fragen (44%) oder die Integration von EbM in die Entscheidungsfindung (35%). Während des Patientenmanagements wurden EbM-Elemente vor allem bei Visiten mit dem „Attending Physician“ (84%), bei den „bedside rounds“ (68%) oder in der Notfallaufnahme (35%) eingesetzt. Allerdings stellte nur die Hälfte der Kliniken typische EbM-Informationsquellen wie „Best Evidence“ oder die

„Cochrane Library“ zur Verfügung. EbM-Ausbildungsangebote für Faculty-Mitglieder fanden nur in 30%-40% der Kliniken statt. Die Evaluation der Programme bewegte sich jedoch meist auf der Befindlichkeits- oder Wissensebene, die Fragen nach Wissenszuwachs, Auswirkung auf Versorgung oder Patientenbefinden wurden nicht beantwortet²³⁹.

4.1.3 EbM in Deutschland

4.1.3.1 Qualifizierungsmöglichkeiten in EbM in Deutschland

In Deutschland hat sich innerhalb von wenigen Jahren ein aktives, nahezu flächendeckendes Kurswesen entwickelt (www.ebm-netzwerk.de). Viele Kursangebote sind echte „Graswurzel“-Aktivitäten, denn im Gegensatz zu Großbritannien, den USA oder Kanada mit finanziell ausgestatteten „Centres for Evidence-based Medicine“ existiert trotz des starken gesundheitspolitischen Akzents auf EbM keine nennenswerte Förderung. Die Kurse orientieren sich in Aufbau und Inhalt stark an den Vorbildern von McMaster²²⁷. In den vergangenen sechs Jahren haben deutschlandweit schätzungsweise 1.500–2000 Teilnehmer einen mehrtägigen Grundkurs durchlaufen. Wegen der Nachfrage nach weiteren Qualifikationsmöglichkeiten wurden EbM-Aufbaukurse entwickelt (Lübeck, Berlin), nach einem ersten Trainingskurs für Tutoren (Train The Trainer 1999; Leitung: Kunz, Raspe, Donner-Banzhoff) sind für Ende 2002 / 2003 weitere Seminare geplant. Mit ihrem Auftrag für die Erstellung eines Curriculums Evidenzbasierte Medizin (s. 4.1.5.4) hat sich die Bundesärztekammer unterstützend hinter diese Form der Weiterqualifikation gestellt.

4.1.3.2 EbM in der institutionalisierten Aus- und Weiterbildung

Die Institutionalisierung von EbM in die medizinische Ausbildung verläuft wesentlich zögerlicher. Nach einer Umfrage der Universität Witten-Herdecke an deutschen medizinischen Fakultäten wird EbM im Studentenunterricht hauptsächlich in den theoretischen Fächer wie Sozialmedizin, Ökologie-Kurs, Biometrie mit einem Zeitkontingent von durchschnittlich ein bis zwei Stunden gelehrt (<http://www.medizinalrat.de/EbM-Ausbildung/ebm-ausbildung.html>). In den praxisorientierten Versorgungsfächern spielt EbM so gut wie keine Rolle. Mit der Verabschiedung der 9. Novelle der

Approbationsordnung für das Medizinstudium ohne die Benennung von EbM als Lehrinhalt ist eine wichtige Gelegenheit ungenutzt geblieben, Studenten bereits zu einem frühen Zeitpunkt der Ausbildung mit der neuen Denkweise vertraut zu machen.

Auch eine systematische Einbindung in die ärztliche Weiterbildung findet derzeit (noch) nicht statt. Im Gegensatz zu den USA, Kanada und Großbritannien gibt es in Deutschland für Assistenzärzte keine strukturierten Weiterbildungsprogramme, in denen sich EbM-Konzepte integrieren ließen. Eine Alternative wäre die Aufnahme von EbM-Kenntnissen in die Weiterbildungsordnungen der einzelnen medizinischen Fachrichtungen. Einige Krankenhäuser haben Pilotprojekte gestartet, um eine entsprechende Infrastruktur aufzubauen (Wuppertal¹¹⁴, Berlin¹¹⁵, Köln, St. Franziskus-Hospital). Leider wurde versäumt, mit einer begleitenden Evaluation die lokalen Erfahrungen systematisch und objektiv zu erfassen.

4.1.4 Was bewirken EbM-Kurse?

4.1.4.1 Ernüchternde Datenlage

Für die Verbreitung von EbM-Fortbildungen unter Ärzten und anderen Gesundheitsberufen finden sich viele plausible Gründe, nicht zuletzt der Einsatz des problemorientierten, selbstbestimmten Lernens als Grundlage des „Lifelong Learning“²⁴⁰. Die Argumente zugunsten einer evidenzbasierten Medizin sind ausreichend ausgetauscht^{195; 223}. Dennoch stellen sich Vertreter wie Kritiker der EbM auch bei dieser Art von Intervention die berechtigte Frage nach der Wirksamkeit: Rechtfertigt das Ergebnis die erheblichen Ressourcen an Zeit, Geld, Personal und das persönliche Engagement von Lernenden wie Lehrenden?

Den fehlenden Beleg des Nutzens nehmen Kritiker und Zweifler^{231; 239; 241} gerne zum Anlass, um das Gesamtkonzept EbM in Frage zu stellen, und auf den ersten Blick scheinen die Daten Recht zu geben: Keiner der vier bisher publizierten Reviews²⁴²⁻²⁴⁵ konnte einen bedeutsamen Effekt bestätigen. Einstimmig üben die Übersichten jedoch Kritik an der methodischen Qualität der eingeschlossenen Studien: schwaches Design mit zu geringer Teilnehmerzahl; unzureichende Beschreibung von Vergleichsgruppen, Settings, (Ko-)Interventionen; Mangel an validierten Instrumenten zur

Leistungserfassung; schlecht gewählte Endpunkte oder viel zu kurze Nachbeobachtungsdauer, um nur einige Defizite zu nennen. Der Cochrane Review von Julie Parkes²⁴⁵ demonstriert das Dilemma, wie eine mangelnde Studienqualität Ergebnisse ohne Aussagekraft produziert: Von über 4000 identifizierten Zitaten aus zwölf Datenbanken wurden schließlich 47 Studien formal bewertet. Nur eine einzige²⁴⁶ erfüllte die a priori festgelegten Einschlussbedingungen.

4.1.4.2 Komplexität der Intervention – Komplexität des Nachweises

Wie definiert man „Nutzen“ in der Bildungsforschung („Educational Research“)? Kirkpatrick^{247; 247} kennt in seiner „Hierarchie der Evaluationsstufen“ vier Ebenen der Verhaltensänderung: die Evaluation der Reaktion (hier: Zufriedenheit mit dem Kurs), die Evaluation des Lernens (hier: erworbene Kenntnisse und Fertigkeiten) und die Evaluation des Verhaltens (Anwendung am Arbeitsplatz). Auswirkung auf die Gesellschaft (hier: die Gesundheit des Patienten) ist die oberste Evaluationsstufe. Je höher die Evaluationsstufe, desto vielschichtiger wird die Verhaltensänderung und entsprechend komplexer und aufwändiger auch das Messverfahren zum Nachweis der Änderungen. In der Regel lassen sich diese nicht von der normalen Umgebung isolieren, sondern laufen parallel zu einer Vielzahl von anderen Ereignissen und Rahmenbedingungen ab, die ebenfalls interagieren²⁴⁸. Damit wird die Messung selbst der wichtigsten Einflussgrößen ausgesprochen schwierig und die Zuordnung ihrer jeweiligen Beiträge zur Wirkung eine Herausforderung. Der mit jeder Stufe steigende Bedarf an Ressourcen erschwert zusätzlich die Durchführung solcher Studien. Gemessen an dieser Hierarchie haben bisher selbst die besten Studien des Reviews die zweite Evaluationsstufe nicht überschritten.

4.1.4.3 Anforderungen an Evaluationsinstrumente

Für einen zuverlässigen Nachweis von Nutzen sind Evaluationsinstrumente mit validen und reliablen psychometrischen Eigenschaften unabdingbar²⁴⁹. Valide Instrumente müssen sorgfältig aufgebaut, pilotiert und getestet und schließlich in einer unabhängigen Population validiert werden, um zu zeigen, dass sie erfassen, was sie zu messen vorgeben^{238; 250}. Entsprechend gehörte die mangelhafte Qualität der Evaluationsinstrumente zu den häufigsten Ausschlussgründen des Parkes-Review. Zur Erfassung

von Verhaltensänderungen wurde vorgeschlagen, OSCEs („Objective Structured Clinical Examinations“) zu entwickeln, wie sie zunehmend in der ärztlichen Ausbildung zur Beurteilung klinischer Fertigkeiten eingesetzt werden²⁴¹. Vermutlich ist diese Methode mit am besten geeignet, um die Anwendung von EbM-Techniken in der Patientenbehandlung angemessen zu erfassen. Angesichts ihrer aufwändigen Entwicklung und zeit- und personalintensiven Testdurchführung wird eine Umsetzung dieser Idee vermutlich noch einige Zeit auf sich warten lassen.

4.1.4.4 Wann wird Evaluation zur Forschung?

Ein möglicher Grund für das schlechte methodische Abschneiden der Studien im Parkes-Review liegt in der Tatsache, dass die meisten Publikationen nicht Forschungsprojekte, sondern die Evaluation von lokalen Aktivitäten beschrieben, die nicht angelegt waren, die methodisch strengen Anforderungen von Forschung zu erfüllen. Die beiden Kategorien „Bildungsforschung“ („educational research“) und „Evaluation von Bildungsmaßnahmen“ werden auch in der Literatur häufig nicht scharf getrennt²⁵¹. In Tabelle 4.1 wurde eine konzeptionelle und methodische Abgrenzung der beiden Kategorien versucht.

Tab. 4.1: Vergleich der Ziele von Evaluationen und didaktischer Forschung^{251; 252}

Evaluation von Bildungsmaßnahmen	Bildungsforschung
Verwendung vergleichbarer Instrumente und Methoden	
<ul style="list-style-type: none"> • Situationspezifische Fragestellungen 	<ul style="list-style-type: none"> • Fragestellungen von allgemeinem Interesse
<p>Ziel: Situationsbezogene Bewertung von Lehrenden und Lernenden zur Verbesserung der Bildungsmaßnahme</p> <ul style="list-style-type: none"> - formative Funktion: Welche Bereiche der Bildungsmaßnahme können verbessert werden? - summative Funktion: Wo lässt sich die Wirksamkeit der Lehre verbessern? 	<p>Ziel: Allgemeingültige Ergebnisse</p>
<ul style="list-style-type: none"> • Ansprüche an Messinstrumente weniger rigoros (z.B. geringere Anforderungen an psychometrische Eigenschaften) 	<ul style="list-style-type: none"> • Hohe Qualitätsanforderungen an Messinstrumente (bei qualitativer wie quantitativer Forschung)
<ul style="list-style-type: none"> • Versuch, die Erfahrungen von Lernenden und Lehrenden zu beschreiben und zu erklären 	<ul style="list-style-type: none"> • Subjektives Erleben von Lernenden / Lehrenden nur sekundäres Forschungsinteresse
<ul style="list-style-type: none"> • Hochwertige Methodik weniger wichtig (möglicherweise hinderlich bei Erfassung des subjektiven Erlebens) 	<ul style="list-style-type: none"> • Hochwertige Methodik wichtig (z.B. Kontrolle für Bias, vollständige Nachbeobachtung, Einsatz komplexer Analysen)
<ul style="list-style-type: none"> • Integration der vorherrschenden Werte bei der Interpretation der Ergebnisse 	<ul style="list-style-type: none"> • Feedback ist kein Ziel Interpretation der Ergebnisse möglichst wertfrei

Die Effective Practice and Organisation of Care (EPOC) – Gruppe der Cochrane Collaboration mit ihrem Themenschwerpunkt „Continuous Medical Education“ hat methodische Kriterien für hochwertige didaktische Studien und Studien über Weiterbildungsmaßnahmen erstellt. Unter anderem werden neben der randomisierten kontrollierten Studie als dem hochwertigsten Design auch kontrollierte Vorher-Nachher-Studien und „unterbrochene Zeitserien“ aufgeführt. Gleichzeitig betont die Gruppe, dass eine randomisierte Zuordnung und insbesondere eine „Verblindung der Intervention“ gerade in der Bildungsforschung oft nicht anwendbar sind und deshalb mit anderen methodischen Techniken gegen Verzerrung und Confounding vorgebeugt werden muss. Diese klar formulierten Ansprüche machen die Abgrenzung zur Evaluation offenkundig.

4.2 Die Berliner EbM-Studie – Nutzwert von EbM-Kursen (eigene Untersuchungen)

Seit 1998 führt unsere Arbeitsgruppe EbM an der Charité (Leitung: Kunz, Fritsche) in Zusammenarbeit mit der Berliner Ärztekammer und der Kaiserin-Friedrich-Stiftung für das Ärztliche Fortbildungswesen Kurse in evidenzbasierter Medizin durch. Um die Inhalte des Kurses für potentielle Teilnehmer transparent zu machen und den Tutoren und Referenten eine Orientierung für die Ausgestaltung der Kleingruppen an die Hand zu geben, erstellten wir in der Arbeitsgruppe einen Gegenstandskatalog, in dem wir das klinisch-epidemiologische und statistische Basiswissen sowie die Grundbegriffe von Literaturrecherche und Informationsmanagement definierten, die wir in unseren EbM-Kursen vermitteln wollten. Nicht zuletzt hofften wir damit auch einen Beitrag zur öffentlichen Diskussion um EbM zu liefern, die offensichtlich von Missverständnissen um die Inhalte, Konzepte und Ziele von EbM geprägt war²⁵³.

4.2.1 Der Berliner Gegenstandskatalog

4.2.1.1 Zielsetzung

Ziel war die Erstellung eines Gegenstandskatalogs zur Definition des klinisch-epidemiologischen und statistischen Basiswissens sowie der Grundbegriffe der Literaturrecherche, die nach unserer Ansicht

für das Verständnis von EbM benötigt werden und in unseren Berliner EbM-Kursen vermittelt werden sollten.

4.2.1.2 Methode der Erstellung

In Anlehnung an Aufbau, Inhalt und Durchführung der EbM-Kurse der McMaster-University und der Kurse des Oxford Centre for Evidence-based Medicine, an denen ich mehrfach als Tutorin teilgenommen hatte und unter Berücksichtigung der Besonderheiten des deutschen Gesundheitswesens wurde der Kurs entwickelt. Für die Definition der Inhalte wurde folgendes Material herangezogen: englisches und kanadisches Kursmaterial, verschiedene Lehrbücher zu EbM²⁵⁴ und der klinischen Epidemiologie¹⁰³, die „Users' Guides“-Serie der McMaster EbM-Working Group²⁵⁵. Für Statistik und Literatursuche wurden klinisch-orientierte Lehrbücher verwendet. Im Gegenstandskatalog wurden vor allem solche Inhalte aufgenommen, die einem praktizierenden Arzt die Logik des Erkenntnisgewinns der klinischen Epidemiologie vermitteln und die für die Entscheidungsfindung in der Patientenversorgung relevant erschienen. Dabei wurden die großen Themengebiete der medizinischen Versorgung abgedeckt: Therapie und Diagnostik, Risikofaktoren und Prognose, Kausalität und Nebenwirkungen; die zugehörigen Studientypen, protektive Maßnahmen zur Verhinderung von Fehlschlüssen und die relevanten Effektmaße. Der Gegenstandskatalog orientierte sich dabei an den Bedürfnissen von Ärzten aus der Patientenversorgung, der Forschung und dem Gesundheitswesen ohne methodische Vorkenntnisse.

Die Themen wurden unabhängig von einem Kliniker / EbM-Tutor aus der Arbeitsgruppe (L.F.) und mir als Klinikerin / klinischer Epidemiologin zusammengestellt, Punkt für Punkt auf ihre Relevanz für die definierten Ziele und Zielgruppen überprüft und im Konsens eingeschlossen.

4.2.1.3 Ergebnisse

Der Berliner Gegenstandskatalog besteht aus drei Säulen: 1) Allgemeine Grundlagen, d.h. Einführung in die Methodik klinisch-experimenteller und Beobachtungsstudien, den Charakteristika von EbM und Grundkenntnissen in Literaturbeschaffung und Statistik; 2) Studientypen aus der Primärforschung für

den Erkenntnisgewinn in den Entscheidungsbereichen des Arztes; 3) Zusammenfassung von Studien als Meta-Analysen oder evidenzbasierte Handlungsempfehlungen (Leitlinien). Eine detaillierte Aufstellung der Inhalte befindet sich im Anhang 4.1.

4.2.1.4 Diskussion

Mit der Erstellung des Gegenstandskatalogs konnten wir demonstrieren, dass sich die Inhalte von EbM eindeutig und transparent benennen lassen. Der Katalog benennt objektive Kriterien, an denen man die Grundlagen einer rationalen Medizin messen und bewerten kann. Durch die Homogenisierung von Kursinhalten zwischen den Ausbildern bietet er eine reproduzierbare Basis für die Durchführung der Schulungsmaßnahmen und eine Grundlage zur Messung der Qualität der eigenen Kurse und des Lernerfolgs der Teilnehmer.

In dem Berliner Gegenstandskatalog werden im Wesentlichen klinisch-epidemiologische Grundkenntnisse für praktizierende Ärzte erfasst. Andere Komponenten der EbM wie Einstellungen, Selbstreflexion und Motivation oder Fertigkeiten wie das Formulieren von strukturierten Fragen aus einem klinischen Problem, das Auffinden der geeigneten Literatur oder die Integration von Patientenvorstellungen in die Entscheidungsfindung werden nicht erfasst. Diese Defizite wurden in dem später entwickelten EbM-Curriculum der Bundesärztekammer berücksichtigt und weiterentwickelt.

4.2.1.5 Schlussfolgerungen

Der Gegenstandskatalog, der Wissen und Fertigkeiten von EbM beschreibt, ist eine wesentliche Voraussetzung für eine rationale und objektive Diskussion um die neue Technik für die Entscheidungsfindung in der Patientenversorgung und zur Steuerung des Gesundheitssystems. Erst durch die explizite Formulierung von Lernzielen wird eine Überprüfung der neu erworbenen Kenntnisse und Fertigkeiten ermöglicht. Mit dem Gegenstandskatalog kann zusätzlich eine

Harmonisierung des steigenden Kursangebots erreicht werden und damit ein qualitätssichernder Einfluss auf Schulungen in EbM ausgeübt werden.

4.2.1.6 Das „Curriculum Evidenzbasierte Medizin“

Angesichts der zunehmenden Bedeutung von EbM für viele Bereiche im Gesundheitswesen hat die Bundesärztekammer im April 2000 die Ärztliche Zentralstelle Qualitätssicherung, Köln und das Deutsche Netzwerk EbM beauftragt, ein Curriculum Evidenzbasierte Medizin zu erstellen. Auf dieser Grundlage sollte es möglich sein, EbM für ein internes Qualitätsmanagement stärker zu nutzen, die Diskussion um Prioritäten und Ressourcen auf eine objektive Grundlage zu stellen, Qualität im Gesundheitssystem besser benennen zu können oder sich mit Ärzten, Patienten und der Öffentlichkeit auf Kriterien für valide Gesundheitsinformationen zu einigen. Dabei stand der Berliner Gegenstandskatalog Pate für das nationale Curriculum Evidenzbasierte Medizin²⁵⁶ (www.ebm-netzwerk.de).

4.2.2 Der „Berliner Fragebogen“: Entwicklung und Validierung eines Instruments zur Messung von EbM-Kenntnissen

Parallel zu der Erstellung des Gegenstandskatalogs wurde unter anderem durch die nationale wie internationale Diskussion um den Nutzen von EbM auch die Notwendigkeit offensichtlich, die Auswirkungen des nicht unerheblichen Aufwands bei den Teilnehmern zu überprüfen. Ein sich ständig wiederholendes Argument in der Kritik war, dass EbM keine Neuerung darstelle und man schon immer nach diesen Prinzipien behandelt habe oder aber dass man für die Patientenversorgung keine EbM benötige. Die Vorlage eines guten Messinstruments erschien uns die beste Antwort auf diese Kommentare und bot uns die Möglichkeit, Kritiker und Skeptiker zur Teilnahme am Test einzuladen.

Unser Ziel war, mit dem Instrument die Fertigkeiten von Ärzten zu messen, Ergebnisse aus der patientenorientierten Forschung zu interpretieren, ein klinisches Problem zu einer spezifischen klinischen Frage umzuwandeln und für deren Beantwortung das beste Studiendesign zu erkennen sowie ihre Fähigkeit zu erfassen, quantitative Informationen aus Studien zur Lösung spezifischer Patientenprobleme zu verwenden.

4.2.2.1 Methodik

4.2.2.1.1 Entwicklung des Berliner EbM-Tests

Anhand der in dem Berliner Gegenstandskatalog²⁵⁷ definierten Kursinhalte entwickelte ein Team aus fünf EbM-Experten - Klinikern, klinischen Epidemiologen und Statistikern mit Erfahrung in der Durchführung von EbM-Kursen - Multiple-Choice-Fragen, in denen klinisch-epidemiologische und statistische Konzepte sowie Wissen aus dem Informationsbereich („informational concepts“) im Kontext von typischen klinischen Situationen (zu Therapie, Diagnose, Prognose, Nebenwirkungen und Übersichtsarbeiten) behandelt wurden. Mit den Fragen, die in klinische Szenarien mit direktem Bezug auf neuere Studien eingebunden waren, sollte das vertiefte Lernen überprüft werden, also die Fertigkeit, Konzepte zu erfassen und sie in anderen Situationen richtig anzuwenden. Falls Berechnungen ein Teil des Lösungswegs waren, wurde der gesamte Lösungsweg angegeben und die Teilnehmer mussten den korrekten Pfad erkennen. Die Fragen wurden nach dem Delphi-Verfahren mehrfach zirkuliert²⁵⁸ und auf eindeutige Wortwahl, eindeutige Lösung, klinische Bedeutung und Wichtigkeit für EbM-Anfänger überprüft. Die Fragebögen wurden von fünf Ärzten ohne EbM-Erfahrung pilotiert und weiter modifiziert. Am Ende lagen zwei Sets à 15 Fragen mit ähnlichen Inhalten, aber in unterschiedlichen klinischen Zusammenhängen für die wiederholten Messungen vor (Berliner Fragebogen s. Anhang 4.2).

4.2.2.1.2 Validierung des Berliner EbM-Tests: Messung von Reliabilität, innere Konsistenz und Diskriminierungsfähigkeit

Bewertung der Schwierigkeit: Um Wissen und Wissenszuwachs sensitiv messen und damit Wissener von den Nicht-Wissern zuverlässig unterscheiden zu können, sollten die Ausgangsfragen nicht zu leicht, aber auch nicht zu schwer sein. Das Endorsement einer Frage (d.h. Prozentsatz richtiger Antwort pro Frage) sollte nicht unter 10% oder über 90% liegen. Die Gleichwertigkeit der beiden Fragensets wurde durch den Intraclass Correlation Coefficient (ICC) der beiden Kontrollgruppen (Experten und Studenten) bestätigt.

Innere Konsistenz (Homogenität des Instruments)^{249; 259}: Mit dem Pearson Korrelationskoeffizient wurde die Beziehung zwischen jedem einzelnen Item und dem Gesamtinstrument (ohne dieses Item)

gemessen (Item-Total-Korrelation²⁶⁰). Items, die zu einer niedrigen Reliabilität beitrugen, wurden identifiziert. Zusätzlich berechneten wir Cronbach's alpha²⁶¹, ein häufig verwendetes Gesamtmaß für innere Konsistenz, bei dem Werte über 0,7 als zufriedenstellend angesehen werden²⁶⁰.

Diskriminierung: Die diskriminierenden Fähigkeiten des Instruments wurden ermittelt, indem Personen ohne Wissen (Studenten nach dem Physikum), Personen mit etwas Wissen (Kursteilnehmer vor dem EbM-Kurs) und solche mit viel Wissen (EbM-Experten) miteinander verglichen wurden, wobei Experten und Studenten beide Fragensets ohne Intervention („EbM-Kurs“) erhielten.

Folgende Vergleiche wurden durchgeführt: 1) Experten vs. interessierte Ärzte vor dem Kurs vs. Studenten; 2) interessierte Ärzte vor dem Kurs: Gruppe A (1999) vs. Gruppe B (2000) vs. Gruppe C (2001). Die Gruppen wurden mit One-way-ANOVA verglichen und bei signifikantem Ergebnis wurde eine post-hoc-Analyse nach Scheffé angeschlossen

4.2.2.2 Ergebnisse

4.2.2.2.1 Beschreibung der Teilnehmer

Insgesamt nahmen 266 Ärzte an den verschiedenen Phasen der Studie teil: 43 EbM-Experten, 20 Studenten und 203 „interessierte Ärzte“ mit gewissen Vorkenntnissen. Tabelle 4.2 bildet das Studiendesign mit den unterschiedlichen Gruppen ab.

Tab. 4.2: Studiendesign: Reihenfolge und Anzahl der von Experten und Studenten vollständig ausgefüllten Fragebögen des Berlin Questionnaire

Kontrollen	Ausgangsfragebogen	Folgefragebogen	Score	Gültige Paare
Experten	Set 1: 30	Set 1: 10	11.7 ± 1.6	20
	Set 2: 13	Set 2: 10	12.5 ± 1.4	
Studenten	Set 1: 10	Set 1: 10	4.0 ± 2.3	20
	Set 2: 10	Set 2: 10	4.5 ± 2.1	

Tab. 4.2 (Fortsetzung): Studiendesign: Reihenfolge und Anzahl der von den Teilnehmern vollständig ausgefüllten Fragebögen des Berlin Questionnaire

Teilnehmer	Ausgangsset (n ausgefüllt) Score ± SD		Folgeset (n ausgefüllt) Score ± SD	Gültige Paare
Kurs A	Set 1 (n = 61) 5.8 ± 2.8	Kurs	Set 2 (n = 60) 9.6 ± 2.4	42
Kurs B	Set 2 (n = 37) 6.9 ± 2.8		Set 1 (n = 46) 10.2 ± 2.7	33
Kurs C	Set 2 (n = 46) 6.6 ± 3.0		Set 1 (n = 70) 10.0 ± 2.2	46
Alle Kurse	n = 143 6.3 ± 2.9		n = 176 9.9 ± 2.4	121

4.2.2.2.2 Validierung des Instruments (s. Anhang 4.2)

4.2.2.2.2.1 Schwierigkeit und innere Konsistenz

Zur Erfassung des Schwierigkeitsgrads der Fragen wurde die Endorsement Rate der einzelnen Fragen berechnet und zwischen den Gruppen mit unterschiedlicher EbM-Expertise verglichen. Die mittlere Endorsement-Rate der Kursteilnehmer lag bei 0.42 (\pm 0.19), die der Experten bei 0.81 (\pm 0.29) und die der Studenten bei 0.29 (\pm 0.43). Somit wurden die Fragen von den Experten als leicht, von den Teilnehmern als relativ schwierig und von den Studenten als sehr schwierig empfunden. Ein Intraclass Correlation Coefficient von 0.96 (95% CI: 0.92 – 0.98) bestätigte die Äquivalenz der beiden Sets. Bis auf eine Frage (Frage 8 in Set II) zeigte sich eine positive Item-Total Correlation für beide Tests. Des weiteren war ein Item von Set I unter der a priori festgelegten Grenze (Frage 9). Cronbach's $\alpha_{\text{Set I}}$ lag bei 0.75 und stieg beim Ausschluss von Frage 9 auf 0.76 an; Cronbach's $\alpha_{\text{Set II}}$ lag bei 0.82 und ohne Frage 8 bei 0.84. Die Fragebögen erfüllten somit die psychometrischen Testeigenschaften für innere Konsistenz.

4.2.2.2.2.2 Diskriminierungsfähigkeit

Das Instrument kann gut zwischen unterschiedlichen Wissensniveaus diskriminieren (Abb. 4.1 – Diskriminierung): Studenten erreichten $4,2 \pm 2,2$ Punkte, interessierte Ärzte $6,3 \pm 2,9$ Punkte und Experten $11,9 \pm 1,6$ Punkte (ANOVA $p < 0.001$), wobei die Unterschiede zwischen jeder dieser Gruppen signifikant waren ($p < 0.01$, Scheffé-Methode). Die Vorkenntnisse der Teilnehmer im Verlauf

der drei Jahre zeigten keinen Unterschied (Gruppe A: $5,8 \pm 2,8$; Gruppe B: $6,9 \pm 2,8$; Gruppe C: $6,6 \pm 3,01$; ANOVA: $p=n.s.$). Damit ist das Instrument sensitiv, um vorhandene Unterschiede in den Kenntnisniveaus wahrzunehmen und stabil, um Gruppen mit vergleichbaren Kenntnissen auch übereinstimmend abzubilden.

Allerdings zeigte sich auch, dass die Kenntnisse der Ärzte nur marginal besser waren als die der Studenten ohne Vorkenntnisse, aber erheblich schlechter als die der Experten.

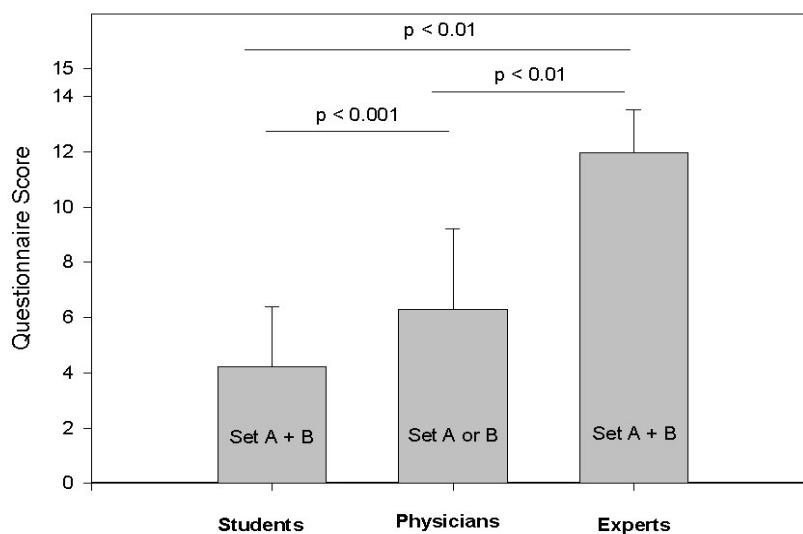


Abb. 4.1: Diskriminierung zwischen Gruppen mit unterschiedlicher EbM-Expertise. Die mittlere Säule repräsentiert den Kenntnisstand der Ärzte vor dem Kurs

4.2.2.2.3 Schlussfolgerungen

Da bekannt ist, dass zwischen Selbstwahrnehmung und objektiver Bewertung eine große Diskrepanz existieren kann²⁶², ist die objektive Auswertung von EbM-Training wiewohl schwierig, so doch unerlässlich. Der Mangel an geeigneten und zuverlässigen Instrumenten ist einer der Gründe. Mit dem Berlin Questionnaire haben wir ein entsprechendes Instrument entwickelt und nachgewiesen²⁶³, dass es zuverlässig zwischen unterschiedlichen Wissensniveaus unterscheiden kann. Diese Ergebnisse widersprechen dem sorgfältig gepflegten Vorurteil mancher Skeptiker, die behaupten, EbM (unter

anderem Namen) schon immer praktiziert zu haben. Ohne theoretische Kenntnisse ist eine evidenzbasierte Praxis schwer vorstellbar.

4.2.3 Die Berliner EbM-Studie

4.2.4 Untersuchung der Wirksamkeit von EbM-Training am Beispiel des Berliner Grundkurses EbM. Eine kontrollierte Vorher-Nachher („Before-After“-Studie (eigene Untersuchungen)

Mit dem neu erstellten Instrument wurde über mehrere Jahre hinweg in den Berliner EbM-Kursen untersucht, inwieweit das intensive Training zu einer Veränderung im Verständnis der Grundkonzepte und der Bewertung von quantitativen Studieninformationen führte. Die Interventionen (die EbM-Kurse) wurden, wie in einer Publikation über die Berliner Kurse²²⁷ beschrieben, nach einem modifizierten McMaster-Format durchgeführt.

4.2.4.1 Methodik

4.2.4.1.1 Messung des Ausbildungseffekts

Insgesamt wurde der Test zwischen 1999 und 2001 dreimal bei dem jährlichen Berliner EbM-Grundkurs eingesetzt. Etwa 60% waren Assistenzärzte, die restlichen meist Fachärzte aus Klinik und Praxis. Die Teilnehmer erhielten den Ausgangstest vier Wochen vor dem Kurs nach Hause zugeschickt, der „Nachher“-Test wurde am letzten Kurstag ausgefüllt. Jedes Jahr wurde die Reihenfolge der Sets geändert (s. Tab. 4.2). Die Kontrollen (Experten und Studenten) erhielten beide Fragebögen in zufälliger Reihenfolge ohne Intervention. Der Kurs dauerte drei Tage und umfasste 22 Stunden Kleingruppenlernen²²⁷ (s. 4.1.2) und fünf Stunden Plenarveranstaltung. In den Kleingruppen, die von einem Methodiker und einem erfahrenen Kliniker geleitet wurden, lag ein besonderer Schwerpunkt auf der Vermittlung typischer EbM-Fertigkeiten wie Identifikation und Priorisierung von Problemen, Formulierung von Fragen, Abwägen verschiedener Behandlungsoptionen und auf dem Erlernen von Grundbegriffen in der kritischen Studienbewertung anhand von Primärstudien im Kontext der Patientenszenarien.

4.2.4.1.2 Zuwachs an Wissen

Der Zuwachs an Wissen und Fertigkeiten wurde mit Hilfe von drei unterschiedlichen Endpunkten überprüft: 1) Veränderung des mittleren Scores nach der Intervention (absoluter Score); 2) unadjustierter relativer Zuwachs; 3) prozentualer Zuwachs, adjustiert für das vorhandene Vorwissen²⁴⁶. Jede richtige Antwort erhielt einen Punkt. Der Wissenszuwachs wurde mit dem t-Test für gepaarte Beobachtungen analysiert.

Sensitivitätsanalyse: Da nicht alle Teilnehmer beide Fragebögen ausfüllten, wurde in einer Sensitivitätsanalyse überprüft, ob dies ein systematisches Problem oder ein zufälliges Phänomen war. Dazu wurden die Ergebnisse der Teilnehmer mit nur einem beantworteten Set mit denen von Teilnehmern mit zwei ausgefüllten Sets verglichen (t-Test für unabhängige Stichproben).

4.2.4.1.3 Studiengröße und Power

Die im Rahmen der Studienplanung durchgeführte Powerberechnung beruhte auf folgenden Annahmen: Alpha-Fehler von 0.05; Beta-Fehler von 0.1, ein mittlerer geschätzter Ausgangsscore von sechs Punkten, mit einer geschätzten Standardabweichung von vier. Ein Wissenszuwachs von mindestens 25% wurde von uns als klinisch relevant angesehen. Damit benötigten wir für unsere Studie 83 Teilnehmer.

4.2.4.2 Ergebnisse

4.2.4.2.1 Beschreibung der Teilnehmer (s. auch 4.2.2.2.1)

An den Kursen nahmen insgesamt 203 „interessierte Ärzte“ mit gewissen Vorkenntnissen teil: 1999 - Gruppe A mit 82 Teilnehmern, 2000 – Gruppe B mit 50 Teilnehmern, 2001 - Gruppe C mit 71 Teilnehmern. Nur 12,3% der Teilnehmer hatten vor dem Kurs in irgendeiner Weise Berührung mit EbM. Beide Fragebögen wurden von 161 der 266 Teilnehmer ausgefüllt. Die Hauptursachen für eine unvollständige Beantwortung beider Sets waren fehlende Abgabe des Vortestsets (n=55), fehlende Teilnahme am Kurs (n=8) oder fehlende Identifizierung (n=38).

4.2.4.2.2 Zugewinn an Wissen und Fertigkeiten

Die Kursteilnehmer zeigten einen hochsignifikanten Zuwachs an Wissen und Fertigkeiten (Abb. 4.2 - Zuwachs an Kenntnissen). Verglichen mit dem Ausgangstest ($6,3 \pm 2,9$ richtige Fragen) beantworteten sie im Durchschnitt zusätzliche 3,6 Fragen richtig ($p < 0.001$; ANOVA). Der Wissenszuwachs war über alle drei Kurse hinweg stabil. Trotzdem blieb das Wissen der Teilnehmer (mittl. Score von $9,9 \pm 2,4$) konstant unter dem der Experten ($11,9 \pm 1,6$; $p < 0.001$). Der nicht adjustierte relative Zuwachs über die drei Kurse lag bei 57%; nach Adjustierung für Unterschiede in den Vorkenntnissen und dem individuellen Potential für Zuwachs an Wissen lag der Anstieg bei 36% ($\pm 46\%$) (Abb. 4.2 - Zuwachs an Kenntnissen).

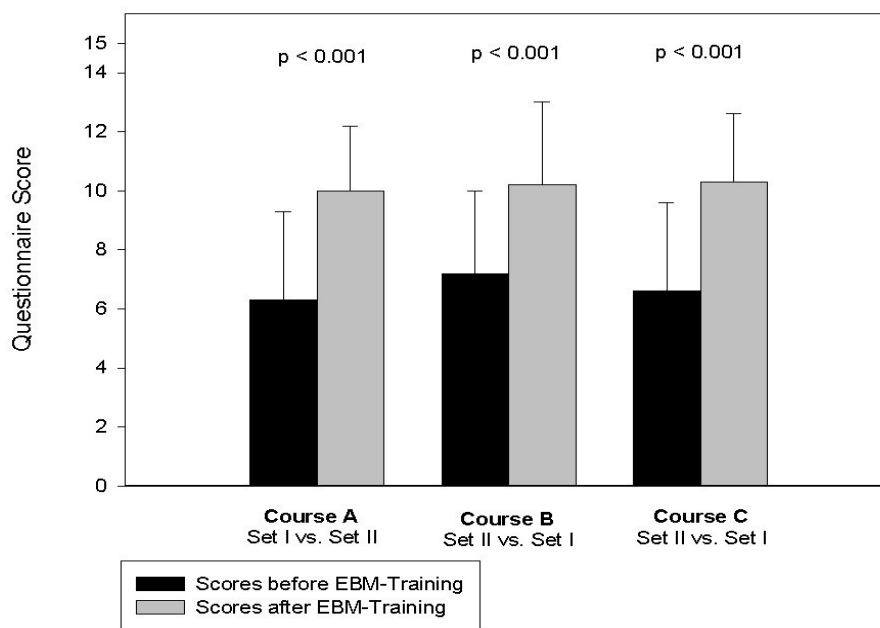


Abb. 4.2: Zuwachs an EbM-Kenntnissen durch Teilnahme an einem EbM-Kurs

Sensitivitätsanalyse: Die Sensitivitätsanalyse zwischen Teilnehmern, die nur ein Set abgeben hatten und denen, die beide Sets abgegeben hatten, zeigte keinen systematischen Unterschied ($p = n.s.$).

4.2.4.3 Diskussion

Im Gegensatz zu den Ergebnissen früherer Reviews zeigt unsere Studie durch den EbM-Kurs einen signifikanten Zuwachs an Wissen und Fertigkeiten in EbM von 57%, den wir auch als didaktisch relevant ansehen. Damit haben wir nachgewiesen, dass 1. die Wirksamkeit von EbM-Kursen objektiv messbar ist und dass 2. unsere Kurse einen solchen Effekt erzielen. Dabei wurde mit unserem validierten Instrument jedoch nur der Kurzzeiteffekt gemessen; über den Langzeiteffekt oder eine Verhaltensänderung kann keine Aussage gemacht werden. Allerdings sind vertiefte Kenntnisse Voraussetzung für Langzeiteffekte oder Verhaltensänderung. Besonders Untersuchungen zu Verhaltensänderungen sollten hoch auf der Forschungsagenda angesiedelt werden. Neben dem Einsatz eines hochwertigen Messinstruments erfüllt die Studie auch die anderen methodischen Kriterien, die die EPOC-Gruppe der Cochrane Collaboration für „Vorher-Nachher“-Studien bei didaktischen Interventionen aufgestellt hat²³³.

Eine Schlüsselfrage für den Nutzen von Ergebnissen ist immer, inwieweit sich die Ergebnisse der Studie auf eine breitere Zielgruppe übertragen lassen. Die freiwillige Teilnahme am Kurs (und damit verbunden eine hohe Motivation), die geschützte Umgebung des Kurses außerhalb des normalen Arbeitsumfelds und die heterogene Mischung von Ärzten aller Ausbildungsstufen und Fachrichtungen wirken stimulierend auf die Lern- und Arbeitsatmosphäre. Die Verankerung des Kurses in den Prinzipien des Erwachsenenlernens (Motivation und Praxisbezug) und des problemorientierten Lernens in Kleingruppen mit aktiver Teilnahme aller Lernenden ist in der Literatur als wichtige Voraussetzung für vertieftes Lernen beschrieben^{195; 231; 263}. Der Bezug zur klinischen Praxis wurde durch die beiden Tutoren sichergestellt, von denen der eine einen methodischen Schwerpunkt, der andere klinische Erfahrung aufwies. Allerdings kann nicht ausgeschlossen werden, dass gerade die Evaluation einen nicht unerheblichen Einfluss auf den Lerneffekt hatte (Hawthorne Effekt oder „Assessment drives Learning“). Wir haben in unserer Studie entsprechende Vorsichtsmaßnahmen eingebaut: freiwillige Teilnahme am Test, Erklärung des experimentellen Charakters des Tests und Betonung seiner Bedeutung als Evaluation und Feedback für die Kursveranstalter und Tutoren, keine Rückmeldung an die Teilnehmer (auch keine Bekanntgabe der richtigen Antworten). Außerdem waren die Tutoren (außer R.Kunz und L.Fritsche) nicht an der Testentwicklung und Durchführung beteiligt.

Es erfolgte eine kontrollierte Abgabe und Rücknahme der Fragebögen, um eine Zirkulation außerhalb der Studie zu verhindern. Trotz dieser neutralisierenden Vorkehrungen kann jedoch nicht ausgeschlossen werden, dass die Evaluation einen motivierenden Effekt auf Teilnehmer und Tutoren hatte. Allerdings wäre das in unseren Augen kein Misskredit, da wir erwarten, dass mehr Veranstalter ihre Kurse routinemäßig evaluieren werden, sobald geeignete Instrumente verfügbar sind und somit die Evaluation zu einem festen Bestandteil des Kurses werden kann.

Weitere Untersuchungen

Als Nächstes stellt sich die Frage, inwieweit der Berliner Fragebogen auch für andere Kurse valide ist und einen Zugewinn an Kenntnissen und Wissen nachweisen kann. Diese Frage wird derzeit in einer internationalen Studie mit einer validierten englischen Übersetzung des Berliner Fragebogens (Berlin Questionnaire) in Großbritannien, Kanada, Italien, in der Schweiz, den USA und in Deutschland untersucht. Dabei werden unterschiedliche Modifikationen des McMaster-Modells, Varianten im curricularen Aufbau oder andere inhaltliche Schwerpunkte erfasst. Von großem Interesse wird zusätzlich der Vergleich der Vorkenntnisse in den einzelnen Ländern sein. Die Datenerhebung soll bis Ende des laufenden Jahres abgeschlossen sein.

Eine zweite Studie ist die Begleitevaluation des nationalen EbM-Curriculums (5.3.4). Dafür soll für jede Kursphase ein eigenes Instrument entwickelt und validiert werden, das einen Effekt über ein breiteres Spektrum an Kursen, Teilnehmern, Themenwahl, Formaten erfassen kann (www.ebm-netzwerk.de).

Forschungsbedarf

Durch die Studie werden eine Vielzahl von weiteren Fragen aufgeworfen: Welche Komponenten in den Kursen bestimmen die Wirksamkeit des Programms? Lassen sich die Ergebnisse auch auf andere Gruppen wie nichtärztliche Gesundheitsberufe, weniger motivierte Teilnehmer oder Studenten übertragen? Wie lassen sich die anderen Kernfertigkeiten von EbM (Entwicklung von Fragen, Fertigkeiten bei der Literatursuche oder differenzierte Anwendung auf den Patienten), die mit zunehmender Nutzung vorbewerteter Evidenz an Bedeutung gewinnen, standardisiert überprüfen?

Sollte man versuchen, auf nationaler oder internationaler Ebene OSCEs („Objective Structured Clinical Examinations“) zu entwickeln? Welche Möglichkeiten gibt es, andere Formen des EbM-Lernens zu bewerten, die von einigen Autoren stark befürwortet werden, wie z.B. Real-Time-EbM-Teaching am Krankenbett²⁶⁴, integrierte EbM-Curricula²³⁸ oder gemischte Fachrichtungen²⁶⁵. Ist der immer wieder geforderte Nachweis der Verhaltensänderung erfassbar? Die Untersuchung dieser Fragen ist aufwändig, aber dringend. Eine erfolgreiche Beantwortung wird von der engen Kollaboration mit lokalen, aber auch mit internationalen Partnern abhängen.

Schlussfolgerungen

Durch EbM-Kurse lassen sich theoretische Grundkenntnisse erlernen, die für eine qualifizierte Anwendung von EbM in der Praxis notwendig sind. Durch das gleichzeitige Kennenlernen von Methoden des selbstbestimmten lebenslangen Lernens wird der Effekt möglicherweise (nachhaltig) verstärkt. Bei der gegenwärtigen gesetzlichen Einführung von evidenzbasierten Leitlinien und evidenzbasierten DMP-Programme wird vernachlässigt, dass ärztliche Kompetenzen in EbM essentielle Voraussetzung für eine evidenzbasierte Versorgung darstellen. Verstärkte Bemühungen um EbM-Fortbildungen könnten die in der Ärzteschaft entstandenen verzerrten bis falschen Vorstellungen über die Konzepte der EbM korrigieren und damit eine bessere Akzeptanz der verschiedenen Versorgungsprogramme erreichen.

5 EbM - Praxisbeispiele und Modellprojekte

Aufbau des Kapitels

Nachdem in den vorherigen Kapiteln die methodischen Grundlagen von EbM (Kapitel 2), ihre Anwendung in ärztlichen Entscheidungssituationen (Kapitel 3) und der Bedarf und Nutzen von EbM-Kursen (Kapitel 4) diskutiert wurde, geht es im letzten Kapitel um die Integration und Integrierbarkeit von evidenzbasierter Medizin in die Versorgung. Diese Frage umfasst sowohl eine evidenzbasierte Medizin im Sackett'schen Sinne, d.h. eine kritische Haltung und spezifische Arbeitsweise des einzelnen Arztes (s. Einleitung und Kapitel 4), als auch die Umsetzung von evidenzbasierten Vorgaben, bei denen der einzelne Arzt mit den Methoden, Arbeitsweisen und Zielen von EbM weniger vertraut sein muss.

Die Literaturübersicht in der Einführung zeigt verschiedene Ansätze, eine evidenzbasierte Versorgung zu etablieren. Die anschließende Studie untersucht, inwieweit man mit spezifischen evidenzbasierten Erläuterungen („EbM-Infos“) zur Krankenhaus-Entlassungsmedikation die Hausärzte motivieren kann, die vorgeschlagene Medikation fortzusetzen (5.2). Diese clusterrandomisierte Studie mit relativ komplexem Design war eingebunden in ein Projekt zur klinikweiten Integration der Methoden und Arbeitsweisen von EbM in einem Krankenhaus der Regelversorgung im Osten Berlins. Als zweites Projekt (5.3) wird ein Pilotprojekt über die Umsetzung eines „evidenzbasierten“ Literaturservice zu aktuellen Fragen aus der Patientenversorgung an einem Universitätsklinikum (der Charité) vorgestellt.

Eigene eingeschlossene Publikationen zum Thema

- Kunz R, Wegscheider K, Zielinski W, Rakowsky N, Donner-Banzhoff N, Neumayer H-H, Müller-Lissner S. Impact of supplemental *EBM-Infos* in discharge letters: The adherence of practitioners to discharge medication. A cluster-randomised controlled trial. (eingereicht)
- Zielinski W, Kunz R. Evidenzbasierte Medizin zwischen Klinik und Praxis. Das Park-Klinik EbM-Projekt – ein Werkstattbericht (Abstrakt und Präsentation). DGSMP. 2000.
- Zielinski W, Kunz R, Rakowsky N, Müller-Lissner S. Niedergelassene Ärzte zur Qualität von Arztbriefen. Ergebnisse einer Befragung aus dem Park-Klinik EbM-Projekt. Projektbaustein EbM-Handlungsempfehlungen. *Gesundh Ökon Qual Manag.* 2002;7:225-231.

- Juche A, Euler U, Bruggenjürgen B, Willich SN, Kunz R. External research service with critical appraisal of the medical literature at a university medical centre. *Qual Saf Health Care* 2002;11:297.
- Euler U, Juche A, Bruggenjürgen B, Kunz R, Willich SN. Implementation of evidence-based medicine in clinical practice - external offers to search for and critically assess medical literature. *Z.Arztl.Fortbild. Qualitätssich.* 2002;96:325-31.

5.1 Evidenzbasierte Medizin - eine Herausforderung für praktizierende Ärzte?

Obwohl die meisten westlichen Gesundheitssysteme inzwischen erkannt haben, dass die Einbindung der EbM in die Versorgung für die Aufrechterhaltung einer hochwertigen Medizin unerlässlich ist¹¹⁷, empfinden viele Ärzte die evidenzbasierte Medizin als erhebliche Herausforderung. Entsprechend stehen sie ihr mit gemischten Gefühlen gegenüber: Die Praxis der evidenzbasierten Medizin erfordert Zeit²⁶⁶ und Engagement: neue spezifische Fertigkeiten müssen erworben werden²⁶⁷, Zugang zu schnellen Informationsdiensten werden benötigt²¹⁶, die oft nicht vorhanden²⁶⁸ oder schwer zugänglich sind²⁶⁹ und nicht immer die gewünschten eindeutigen Entscheidungshilfen bieten^{270; 271}. Insbesondere Allgemeinärzte klagen, dass die aufgefundenen Studien die Komplexität ihres Versorgungsalltags nicht widerspiegeln²⁷². Vielfach sehen sie die Arbeitsweise der hausärztlichen Versorgung einer quantitativen Forschung nicht zugänglich²⁶⁶ und finden ihre Arbeitswelt in der evidenzbasierten Medizin nicht abgebildet. Insbesondere, wenn die Erwartungen des Patienten nicht mit der Evidenzlage übereinstimmen, erleben sie die EbM als unnötige und unerwünschte Störung der Beziehung zu ihrem Patienten^{266; 273}.

Auf der anderen Seite werden Ärzte während der Patientenkontakte regelmäßig mit ihren eigenen Wissenslücken konfrontiert^{274; 275}. So kamen im Durchschnitt auf drei Patienten, die Assistenzärzte während ihrer Sprechstunde betreuten, zwei klinisch relevante Fragen, von denen (trotz Studiensetting) weniger als 30% nachverfolgt wurden²⁷⁶. Ähnliche Ergebnisse wurden für Hausärzte beobachtet²¹⁶. Das Ausmaß des Informationsdefizits wurde inzwischen in einer Vielzahl von Studien ausreichend dokumentiert und demonstriert die Schwierigkeiten der Ärzte, trotz Informationsüberfluss^{277; 278} mit den Entwicklungen selbst im eigenen Fachgebiet Schritt zu halten^{214;279}. Auswirkungen auf die Qualität der Versorgung bleiben nicht aus^{186; 280}. Die

evidenzbasierte Medizin ist eine Herausforderung an die Ärzte. Ist sie eine unzumutbare Belastung? Scheitert sie an der Umsetzbarkeit?

5.1.1 Praxisprojekte zur Implementierung von EbM in die Versorgung

Die Einführung einer evidenzbasierten Medizin in die Versorgung ist anspruchsvoll. Sie fordert von den Ärzten Veränderungen in ihren über Jahre erworbenen Einstellungen^{195; 256}, Bereitschaft zur Reflexion der eigenen Arbeitsweise und Hinterfragung scheinbar funktionierender Versorgungsabläufe. Veränderungen sind jedoch nur schwer zu erzielen. Wie die Erfahrungen der Leitlinienbewegung eindeutig belegen, reicht die alleinige Bereitstellung von Informationen nicht aus²⁸¹. Vielmehr muss die Forderung nach einer evidenzbasierten Gesundheitsversorgung, wie sie in Deutschland im Sozialgesetzbuch V § 137 festgeschrieben wurde, durch „evidenzbasierte“ Implementierungsstrategien ergänzt werden^{280; 282}. Die Frage, wie die gewünschten Veränderungen erfolgreich herbeigeführt und woran der Erfolg gemessen werden sollen, bleibt in der deutschen Debatte um die Integration der evidenzbasierten Medizin in das Gesundheitswesen jedoch weitgehend ausgespart. Welche Rahmenbedingungen werden benötigt? Welche Faktoren und Prozesse fördern die Umsetzung in die Praxis, welche verhindern sie ?

Zwei Projekte aus den Anfangszeiten von EbM in Großbritannien beschäftigten sich vor allem mit der Umsetzbarkeit der Sackett'schen EbM-Techniken (s. 4.1.2 und 4.1.3.1) in der stationären und ambulanten Routineversorgung. Dabei analysierten sie vor allem die vorhandenen Rahmenbedingungen und Prozessabläufe, um typische unterstützende und hinderliche Faktoren zu identifizieren („Prozessorientierte EbM-Projekte“: FRONTLINE- und PACE-Projekt). Andere Projekte orientierten sich stärker an der gezielten Beseitigung von Fehlversorgung bei einem definierten Krankheitsbild mit guter Evidenzlage. In randomisierten Studien überprüften sie punktuell, ob durch Änderungen von Prozessabläufen der Routineversorgung oder durch intensivere Betreuung die Patienten evidenzgerechter versorgt werden können (EbM-Versorgungsstudien).

5.1.1.1 Prozessorientierte EbM-Projekte: FRONTLINE-Projekt und PACE-Projekt

Im „FRONTLINE ebm-Project“²⁸³ wurde untersucht, inwieweit sich eine Patientenversorgung nach den Methoden der EbM in der Versorgungsrealität britischer Krankenhäuser der Regelversorgung (in denen der größte Teil der stationären Patientenversorgung stattfindet) umsetzen lässt. Dazu wurden die strukturellen Rahmenbedingungen, die inhaltlichen Voraussetzungen (insbesondere Kenntnisstand der Ärzte) und die Akzeptanz von spezifischen EbM-Weiterbildungen erfasst. An dem Projekt beteiligten sich 14 Krankenhäuser mit 20 Teams („firms“) aus verschiedenen Fachrichtungen und mit unterschiedlichem Weiterbildungsstand. Im Gegensatz zu den oft kritischen bis ablehnenden Stellungnahmen in der Literatur^{266; 272} fanden die Ärzte die Ideen und Konzepte der EbM interessant und für die Versorgung ihrer Patienten relevant. Allerdings waren das Bewusstsein für vorhandene Informationsdefizite gering ausgeprägt und die klinisch-epidemiologischen Basiskenntnisse stark ausbaufähig. Eklatante Mängel zeigten sich bei strukturellen Voraussetzungen wie der Computerausstattung auf den Stationen, der Verfügbarkeit von relevanten Informationen („Most wards are still information-free zones as far as statistical research findings are concerned“ A.Donald²⁸³) oder dem Zugriff auf Literatur, wobei sich insbesondere bei der Ausstattung der Bibliotheken ein beträchtlicher Nachrüstungsbedarf zeigte.

Das zweite EbM-Projekt über Prozessabläufe und Rahmenbedingungen war das PACE-Projekt („Promoting Action on Clinical Effectiveness: to understand how to implement change in clinical practice“)²⁸⁴, das aus 16 Einzelprojekten in heterogenen Versorgungssituationen (stationäre / hausärztliche Versorgung, somatische Medizin / Psychiatrie, ambulante / stationäre Pflege) bestand. In Bereichen mit lokaler Unter- oder Fehlversorgung (z.B. „*Prävention von Schlaganfall – Hypertonie bei den Älteren*“, „*Eradikation von H.pylori beim Management von Dyspepsie*“, „*Post-operative Schmerzkontrolle*“) sollten die einzelnen Projektleiter eigenständig Ansätze erproben, wie sich der Einsatz diagnostischer oder therapeutischer Interventionen mit robuster Evidenzbasis in der Routineversorgung steigern lässt. Die Ergebnisse wurden gemeinsam ausgewertet und sich wiederholende Problemfelder und Lösungsansätze näher beleuchtet.

Typische Schwierigkeiten waren ablehnende Einstellungen gegenüber Leitlinien, Änderungen tradierter Rollen und Zuständigkeiten oder die Umsetzbarkeit der Empfehlungen (z.B. Sicherstellung

zusätzlicher Endoskopieservices). Erfolgreiche Strategien waren die Einbindung von Meinungsbildnern, der gezielte Abbau von Vorurteilen zwischen Chefärzten im Krankenhaus und Hausärzten oder die Entwicklung einer Mitverantwortung für die angestrebten Veränderungen („Ownership“) unter den beteiligten Ärzten. Eine gesicherte Evidenzlage wirkte als Autorität. Unter diesen Bedingungen war es möglich, etablierte Verhaltensweisen zu ändern und den Einsatz dieser Maßnahmen zu steigern. Zusätzlich war ein engagiertes Projektmanagement für den Erfolg der Projektentwicklung und -umsetzung entscheidend. Die Prozesse der Veränderung liefen in der Regel nicht geradlinig und „nach Plan“, sondern dynamisch, scheinbar unorganisiert und oft wenig voraussehbar. Ausnahmslos bestätigten die Projekte den großen Aufwand an Personal, Zeit, Ressourcen und den erheblichen persönlichen Einsatz der Ärzte, um den angestoßenen Prozess auf dem Laufenden zu halten.

5.1.1.2 EbM-Versorgungsstudien

Bei den Versorgungsstudien über EbM mit experimentellem Design standen nicht die Prozesse, sondern die verbesserte Nutzung definierter „evidenzbasierter“ Interventionen (d.h. Therapien mit gut belegter Wirksamkeit) in Bereichen mit Fehlversorgung im Vordergrund. Durch Änderung der herkömmlichen Versorgungsprozesse wurden neue spezifische Versorgungsmuster erprobt.

Die Ergebnisse sind nicht eindeutig: Zwei RCTs untersuchten die Auswirkung von patientenzentrierten Maßnahmen auf die Sekundärprävention von KHK: In der ersten Studie sollte ein Erinnerungsschreiben²⁸⁵, eine eigentlich effektive Maßnahme zur Verhaltensänderung²²⁰, die Verordnungsrate von Aspirin, β -Blockern oder Statinen bei Post-Infarktpatienten verbessern. Die 2. Studie zielte mit einem Schulungsprogramm für Infarktpatienten auf eine Verbesserung der Intermediärendpunkte Blutdruck, Rauchverhalten oder Cholesterinspiegel²⁸⁶. Trotz Unterversorgung kam es in keinem Programm zu einer besseren Nutzung der Sekundärprophylaxe.

Erfolgreicher waren dagegen RCTs, in denen die Ärzte relevante „EbM-Informationen“ für patientenspezifische Probleme zum Zeitpunkt der Entscheidungsfindung erhielten: Pädiater, die während der Ausstellung eines Rezepts für die Behandlung von Otitis media an die Evidenzlage erinnert wurden, verordneten signifikant weniger Antibiotika²⁸⁷. Durch die Verknüpfung von

Praxisdaten mit der externen Evidenz aus entsprechenden Studien steigerte sich in der Interventionsgruppe der Einsatz der Peritonealdialyse von 2,4% auf 15,3%²⁸⁸. Eine von medizinischem Fachpersonal geführte „Sprechstunde“ zur Sekundärprävention der KHK führte zu einer signifikanten Senkung der Risikofaktoren und verstärktem Einsatz der protektiven Medikamente²⁸⁹. In allen Situationen wurden die „EbM-Interventionen“ zum Zeitpunkt des Patientenkontakts in die Versorgung integriert, so dass der Arbeitsablauf nicht unterbrochen werden musste. Vielversprechend ist eine vor kurzem publizierte systematische Übersicht über RCTs zu sogenannten „Disease-Management-Programmen“ (DMP) bei KHK-Patienten zur Sekundärprävention²⁹⁰, bei denen die Nutzung präventiver Interventionen mit ganz unterschiedlichen Ansätzen strukturiert und organisiert wurde. In den Programmen kam es zu einer signifikanten Verbesserung der Prozessdaten (Verordnung von Lipiden, β -Blockern, Senkung der Risikofaktoren) verglichen mit der jeweiligen „normalen Versorgung“. Die kurze Nachbeobachtungsphase erlaubte jedoch keine Aussage über den Einfluss auf die Mortalität. In dieser Übersicht zeigte sich, dass in den Programmen gesicherte Maßnahmen verstärkt genutzt werden, wenn parallel zur Information der Ärzte auch die organisatorischen Prozesse gezielt modifiziert werden. Allerdings lieferten viele Studien, die solche und ähnliche Fragestellung untersuchten, aufgrund der eingesetzten schwachen Designs nur wenig aussagekräftige Ergebnisse. Weitere Untersuchungen werden benötigt.

5.1.1.3 Schlussfolgerungen

Obwohl Prinzip und Grundlagen von EbM logisch, einfach und selbsterklärend erscheinen, ist die Integration in die Versorgung wesentlich komplexer als weithin angenommen. Die Studien bestätigen einheitlich, dass fehlendes Wissen und Fertigkeiten, ablehnende Einstellungen, technische Hürden und hinderliche Prozessabläufe überwunden werden müssen. Verschiedene Implementierungsstrategien scheinen dabei besser geeignet als andere, wobei zu berücksichtigen ist, dass Strategien, die in einer Versorgungssituation erfolgreich sind²²⁰, in anderen Situationen²⁸⁵ nicht automatisch den gleichen Effekt herbeiführen. Kontextbezug und engagiertes Projektmanagement sind weitere wichtige Faktoren, die erheblich zum Projekterfolg beitragen, aber viel schwieriger zu messen und zu

vergleichen sind. Alle Programme bestätigen jedoch einheitlich, dass für einen erfolgreichen Prozess der Veränderung an mehreren Stellen des Versorgungsprozesses gleichzeitig angesetzt werden muss. Die eindeutigen Aussagen über die Voraussetzungen für eine erfolgreiche Einführung von EbM haben wir im EbM-Projekt der Park-Klinik Weissensee aufgegriffen und ein multimodulares Konzept entworfen und umgesetzt. Dieses Projekt soll im Folgenden vorgestellt werden.

5.2 Das Park-Klinik Weissensee - EbM-Projekt (eigene Untersuchungen)

Mit dem Park-Klinik Weissensee (PKW)-EbM-Projekt, einem Modellprojekt der Park-Klinik Weissensee (PKW) Berlin, der Ärztekammer Berlin und der Technikerkrankenkasse Hamburg wurde in Deutschland erstmalig der Versuch unternommen, klinikweit die Methoden und Arbeitsweisen von EbM einzuführen und auf verschiedenen Ebenen in die routinemäßige Patientenversorgung zu integrieren. Die Park-Klinik Weissensee ist ein Krankenhaus der Regelversorgung im Osten Berlins mit privatem Träger und akademisches Lehrkrankenhaus der Charité. Die Klinik verfügt über neun Fachabteilungen und 350 Planbetten. Sie besitzt langjährige Erfahrung im Qualitätsmanagement (KTQ), was eine gute Voraussetzungen für die Umsetzung der notwendigen Prozessveränderungen darstellt.

5.2.1 Das Gesamtprojekt

Das Projekt bestand aus vier Teilprojekten, die in unterschiedlicher Verantwortung und Organisation umgesetzt wurden¹¹⁵:

1. ein intensives Schulungsprogramm in den Grundlagen der EbM und den Techniken der Literaturrecherche durch Teilnahme am Berliner Grund- und Aufbaukurs (s. 4.2) und an hausinternen Schulungen;
2. eine Untersuchung zur Sicherstellung einer wissenschaftlich fundierten („evidenzbasierten“) Versorgung an der stationären / ambulanten Schnittstelle; im Rahmen einer clusterrandomisierten kontrollierten Studie wurden die Auswirkungen von EbM-gestützten

Handlungsempfehlungen in Entlassungsbriefen auf die Fortführung der Behandlung in der ambulanten Versorgung untersucht;

3. ein Leitlinienkolloquium mit der Entwicklung und Implementierung von klinikinternen, z.T. abteilungsübergreifenden Versorgungsleitlinien und klinischen Behandlungspfaden (z.B. präoperatives Management von Patienten mit Diabetes mellitus);
4. Informationsveranstaltungen und Fortbildungsangebote über EbM und das EbM-Projekt für die niedergelassenen Ärzte im Einzugsbereich der Park-Klinik Weissensee.

Zusätzlich wurden die notwendigen Ressourcen wie Datenbanken mit vorbewerteter Evidenz, Cochrane Library, Best Evidence, Clinical Evidence, hausinterne Leitlinien und arbeitsplatznahe Zugriffsmöglichkeiten bereitgestellt.

Das Gesamtkonzept wurde maßgeblich auf der Grundlage der Vorschläge und Entwürfe der Autorin entwickelt, die Verantwortung und Realisierung der einzelnen Teilprojekte wurden aufgeteilt. Die Projektleitung für die clusterrandomisierte Studie lag in meiner Verantwortung; zusätzlich leistete ich wesentliche Beiträge bei der Erstellung des Schulungsprogramms für die Mitarbeiter (Teilprojekt 1) und der Information und Einbeziehung der niedergelassenen Ärzte (Teilprojekt 4). Der Klinikleitung bzw. dem Qualitätsmanagement der PKW oblag die Verantwortung für die Teilprojekte 1, 3 und 4. Die clusterrandomisierte Studie startete im Herbst 1999 und wurde im Herbst 2001 abgeschlossen. Die drei Teilprojekte laufen weiter.

Mit dem klinikweiten multimodularen Ansatz und der Unterstützung von Verwaltung und ärztlicher Leitung gibt es nach unseren Kenntnissen deutschlandweit kein vergleichbares Projekt. Bedauerlicherweise wurde von einer systematischen Begleitevaluation des gesamten Projekts Abstand genommen, wodurch viele dringende Fragen für die Umsetzbarkeit von EbM-Methoden in Kliniken der Regelversorgung unbeantwortet bleiben. Auswirkungen dieses aufwändigen und intensiven Projekts auf Krankenhausprozesse, das ärztliche Verhalten oder auf die Patienten können dadurch nicht adäquat gemessen und bewertet werden. Damit blieb eine Gelegenheit ungenutzt, die Wirksamkeit der Implementierung von EbM in dem engen Zeitpunkt der frühen Verbreitung zu

untersuchen, in dem die Störeffekte von Ko-Interventionen, einer diffusen Verbreitung von EbM und allgemeine Lerneffekte noch gering ausgeprägt sind.

Die folgende Beschreibung beschränkt sich auf das von mir geleitete Teilprojekt 2: EbM-gestützte Handlungsempfehlungen („*EbM-Infos*“) in Arztbriefen.

5.2.2 EbM-gestützte Handlungsempfehlungen in Arztbriefen

5.2.2.1 Hintergrund

Je nach Schweregrad der Erkrankung, dem Vorhandensein lokaler Expertise oder den vorgehaltenen diagnostischen und therapeutischen Möglichkeiten werden Patienten ambulant, teilstationär oder vollstationär behandelt. Dabei erfolgt im deutschen Gesundheitssystem die Versorgung ambulanter Patienten nahezu ausschließlich von niedergelassenen Ärzten, die Versorgung am Krankenhaus ist bis auf wenige, zahlenmäßig unbedeutende Ausnahmen (universitäre Spezialambulanzen, Belegärzte) ausschließlich stationären Patienten zugänglich. Es besteht also eine strikte Trennung der beiden Sektoren. Die daraus resultierenden Konsequenzen für Kommunikationsprozesse und Therapieentscheidungen wurden kürzlich in einer Studie des norddeutschen Public Health Verbunds zusammengefasst²⁹¹. Dabei zeigt sich, dass angesichts der Diskontinuität der Versorgung an der Schnittstelle „stationär / ambulant“ erhebliche Kommunikationsdefizite zwischen den stationär behandelnden und den ambulant weiterbetreuenden Ärzten bestehen. Häufig herrscht Unklarheit über die Fragestellungen des ambulanten Arztes oder über Vorbefunde des Patienten; der niedergelassene Kollege erhält nur unzureichende Informationen über den stationären Verlauf, über die Ergebnisse bzw. Erfolge von diagnostischen und therapeutischen Maßnahmen und rudimentäre Empfehlungen für die Weiterbehandlung²⁹². Ähnliche Probleme werden auch aus anderen Ländern berichtet^{293; 294}. So überrascht es nicht, wenn Behandlungen, die während des Krankenhausaufenthalts begonnen wurden, vom niedergelassenen Arzt geändert oder abgesetzt werden. Die beschriebenen Verhaltensweisen sind den Ärzten in beiden Sektoren gut bekannt²⁹¹, der Forschungsstand über das Ausmaß einer fehlenden Akzeptanz von Behandlungsempfehlungen ist angesichts seiner Bedeutung für die

Versorgungsqualität erstaunlich gering. Eine neuere prospektive Studie aus Deutschland berichtet über eine Fortsetzungsrate von nur 60% unter den Allgemeinmedizinern²⁹⁵.

Um die Kontinuität an der Versorgungsschnittstelle stationär / ambulant zu verbessern, wurde nach Optionen gesucht, die Angemessenheit der angesetzten Behandlungen stärker zum Ausdruck zu bringen. Untersuchungen haben wiederholt gezeigt, dass wissenschaftlich gut belegte Maßnahmen (z.B. PACE-Projekt) eine eigene Autorität besitzen, die von Ärzten in der Regel anerkannt wird. Die alleinige Verfügbarkeit von guter Evidenz garantiert jedoch keine konsequente Umsetzung. Unterstützende Maßnahmen wie z.B. die Bereitstellung „evidenzbasierter“ Behandlungsempfehlungen zum Zeitpunkt der Entscheidungsfindung haben sich bewährt^{287; 288}.

In der vorliegenden Studie zur Verbesserung der Versorgungsqualität an der stationär-ambulanten Schnittstelle soll deshalb untersucht werden, ob mit einer kurzen spezifischen Erläuterung der Evidenzbasis im Arztbrief die Hausärzte motiviert werden, im Krankenhaus neu angesetzte Langzeitbehandlungen in der ambulanten Versorgung fortzusetzen.

5.2.2.2 Fragestellung der Studie

Primäre Fragestellung: Führen kurze Erläuterungen (sog. „*EbM-Infos*“) in Krankenhausentlassungsbriefen zu Behandlungsempfehlungen mit hohem Evidenzgrad zu einer Reduktion der Absetzrate durch die Hausärzte, verglichen mit Entlassungsbriefen ohne Erläuterung?

Sekundäre Fragestellung: Welche Faktoren sind dabei für die Entscheidungen der Ärzte ausschlaggebend? Wie bewerten Ärzte den Nutzen dieser zusätzlichen Informationen ?

5.2.2.3 Studiendesign und Methodik

5.2.2.3.1 Setting der Studie

Die Studie „*EbM-gestützte Handlungsempfehlungen*“ wurde in der Inneren Abteilung der PKW (Ltg. Prof. Müller-Lissner) und dem zugehörigen Einzugsgebiet durchgeführt. Alle Mitarbeiter hatten Schulungen in Recherchetechniken durchlaufen, ein Großteil hatte am Berliner EbM-Grundkurs (4.2),

einige auch an dem Aufbaukurs teilgenommen. Diskussionen über die Evidenzbasis angesetzter Behandlungen sind integrierter Bestandteil der ärztlichen Entscheidungsfindung während der Visiten.

5.2.2.3.2 Design und Clusterrandomisierung

Die Studie wurde mit einem clusterrandomisierten kontrollierten Design durchgeführt. Die Randomisierung in Clustern ist erforderlich, wenn Interventionen in organisatorischen oder administrativen Einheiten (z.B. Schulen, Kindergärten, Krankenhäusern oder Arztpraxen) stattfinden, deren Mitglieder Gemeinsamkeiten („Gruppeneffekte“) aufweisen, und die Gefahr besteht, dass durch Gruppeneffekte die Endpunkte der Individuen innerhalb der Cluster einander ähnlicher sind als die Endpunkte zwischen verschiedenen Clustern. Diese Abhängigkeit muss in Design wie Analyse der Studie berücksichtigt werden. Für unsere Studie bedeutete dies die Randomisierung auf Praxisebene, wengleich die Interventionsebene der einzelne Arztbrief, d.h. der Patient war. Denn jeder Arzt konnte potentiell mehrere Studienpatienten behandeln, und in Gruppenpraxen konnten möglicherweise mehrere Ärzte den gleichen Patienten betreuen. Jeder Arzt hat eigene Behandlungsvorstellungen, die er bei seinen Patienten anwendet. In jeder Gruppenpraxis gibt es möglicherweise zwischen den Partnern implizite oder explizit festgelegte gemeinsame Behandlungsstrategien, alles Gruppeneffekte, die sich auf das Absetzverhalten auswirken können. Ohne Clusterrandomisierung besteht in Gruppenpraxen zusätzlich die Gefahr von Übertragungseffekten (Kontamination), wenn ein Patient mit Handlungsempfehlung von einem Arzt der Kontrollgruppe betreut wird (Abb. 5.1). Durch die Anwendung der Clusterrandomisierung wird sichergestellt, dass trotz dieser unvermeidlichen Gruppeneffekte die für die statistische Analyse erforderliche Unabhängigkeit der „sampling units“ unterstellt werden kann. Während sich die meisten clusterrandomisierten Studien auf zwei Ebenen - der Interventionsebene (hier: der einzelne Arztbrief / Patient) und der Evaluationsebene (hier: die (Gemeinschafts-) Praxis)²⁹⁶ - abspielen, existiert in unserer Studie durch die Gemeinschaftspraxen noch eine intermediäre Ebene der Ärzte, von denen jeder einen oder mehrere Patienten aus der Studie betreuen konnte (s.Abb.5.1).

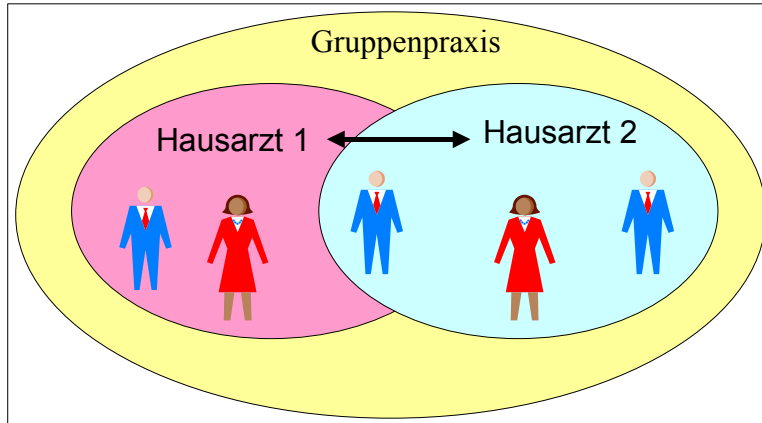


Abb. 5.1. Randomisierung von Arztpraxen (= Cluster)
 In einer Gruppenpraxis (großes Oval) können die Patienten (Figuren) zwischen den einzelnen Hausärzten (kleine Ovale) nicht sicher getrennt werden (Problem der Kontamination). Zusätzlich muß die Existenz von Gruppeneffekten zwischen den Ärzten (z.B. ähnliches Ordnungsverhalten, praxisinterne Absprachen) angenommen werden.

Das Clusterdesign zieht methodische Anforderungen an die Studie nach sich³⁵: Bei der Auswertung können die auf der Analyse von einzelnen Patienten beruhenden Standardverfahren nicht eingesetzt werden, da die Annahme einer Unabhängigkeit der Antworten der einzelnen Individuen durch den vermuteten Gruppeneffekt nicht erfüllt wird. Durch den Clustereffekt sind für die Berechnung der Sample Size spezielle Methoden erforderlich²⁹⁷. Die Power einer clusterrandomisierten Studie ist im allgemeinen geringer als die einer individuell randomisierten Studie und der Stichprobenumfang muss entsprechend höher angesetzt werden. Ferner muss eine hinreichende Anzahl an genügend großen Clustern rekrutiert werden.

Bereits im Design müssen potentielle (starke) Confounder, die innerhalb der Cluster wirksam werden können, berücksichtigt werden, um diese a priori (durch Stratifizierung oder Matching) oder in der Analyse durch Adjustierung zu kontrollieren. Alle diese Aspekte sind in der Planung, Durchführung und Auswertung dieser Studie berücksichtigt worden.

Durchführung: Grundgesamtheit der Studie waren alle Praxen, die 1999 mindestens einen Patienten in die PKW eingewiesen hatten. Diese wurden a priori randomisiert und die Randomzuteilung wurde

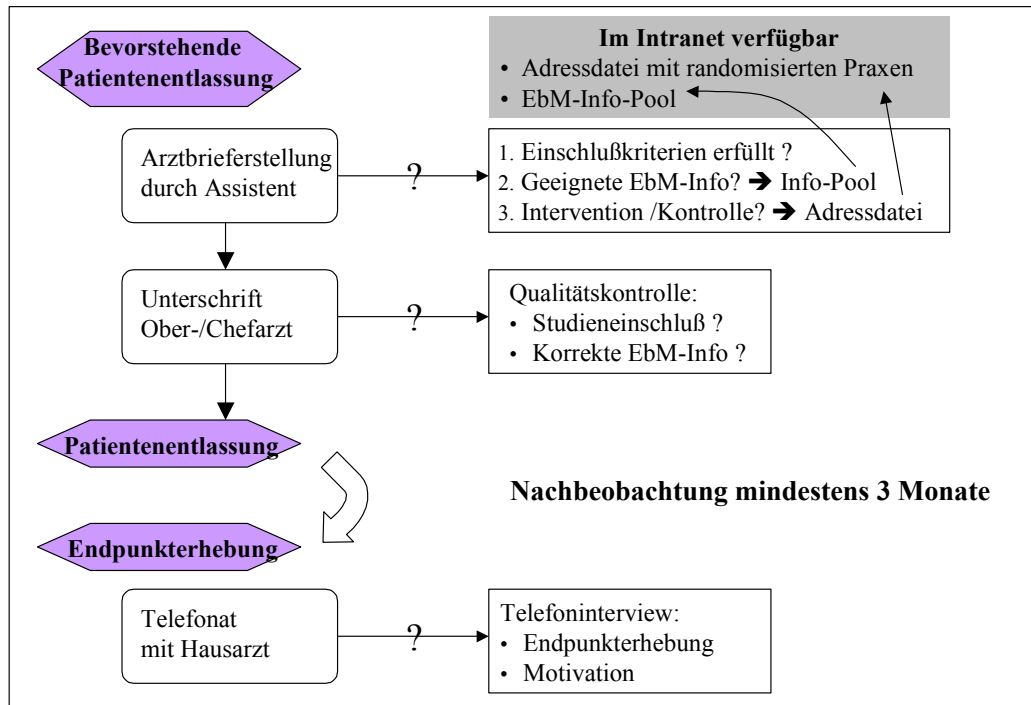


Abb. 5.2: Ablauf von Rekrutierung und Endpunkterhebung

in die Adressdatei eingetragen. Für Patienten aus nichtrandomisierten Praxen (d.h.1999 war von diesen Praxen keine Einweisung erfolgt) wurde eine Nachrandomisierungsliste im Sekretariat hinterlegt. Bei der Randomisierung der Grundgesamtheit war klar, dass nur ein Teil der randomisierten Praxen auch in die Studie rekrutiert würde. Eine Praxisrandomisierung zum Zeitpunkt der Rekrutierung während der Arztbriefschreibung, ggf. auch spät abends, war in dem Routineklinikbetrieb nicht umsetzbar. Beim Verfassen des Arztbriefs (die Patienten erhielten den Arztbrief zum Zeitpunkt der Entlassung) wurden die Patienten für die Studie gescreent. **Einschlusskriterien** waren: Beginn einer Langzeitmedikation, für die im „EbM-Info“-Pool (s.u.) eine entsprechende Empfehlung zur Verfügung stand. **Ausschlusskriterien** waren: Lebenserwartung < 6 Monate, Überweisung in ein anderes Krankenhaus oder Hausarzt außerhalb des Einzugsbereichs der PKW. Die Rekrutierung erfolgte von 5/2000 – 12/ 2000.

5.2.2.3.3 „EbM-Info“-Pool

Für die Studie wurde von den EbM-geschulten Ärzten und Oberärzten der Abteilung ein Pool mit „evidenzbasierten“ Behandlungsempfehlungen („*EbM-Infos*“) erstellt. „*EbM-Infos*“ sind in einem Satz zusammengefasste Therapiestudien (randomisierte, kontrollierte Studien oder Meta-Analysen mit höchster Evidenzstufe – „Level of Evidence I“, <http://cebm.jr2.ox.ac.uk/>) zu häufigen internistischen Erkrankungen, die im Krankenhaus behandelt werden (Abb.5.3). Grundlage für die meisten Textbausteine waren vorbewertete Studien aus Best Evidence²³⁴, der Cochrane Library²³³ und Clinical Evidence²³⁵. Die Ergebnisse wurden mit der entsprechenden Evidenzstufe (Level of Evidence, LoE) und der Number Needed to Treat (NNT) angegeben. Jede „*EbM-Info*“ erhielt eine einmalige Identifikationsnummer. Die erstellten „*EbM-Infos*“ wurden vor Studienbeginn von dem Abteilungsleiter, Prof. Müller-Lissner und der Autorin auf Richtigkeit, innere Konsistenz und Eignung für die Studie überprüft und für die Studie ausgewählt. Für die Studie bestand der Pool aus 135 „*EbM-Infos*“ aus 15 vorab definierten diagnostischen Gruppen (z.B. Hypertonie; Vorhofflimmern; chronische Bronchitis; Fettstoffwechselstörungen).

Abb. 5.3.: Beispiel für eine „*EbM-Info*“

ACE-Hemmer bei Herzinsuffizienz

Nach Myokardinfarkt verbessern ACE-Hemmer die Prognose einer Herzinsuffizienz durch eingeschränkte Ejektionsfraktion (z.B. Trandolapril NNT_{3 Jahre}: 11).

Je schlechter die linksventrikuläre Funktion, desto größer der Nutzen.

Lancet 1999; 354:9-12

5.2.2.3.4 Die Intervention

Die Hausärzte der Interventionsgruppe erhielten als Ergänzung zu dem üblichen Entlassungsbrief mit den Medikamenten an einer gut sichtbaren Stelle die „*EbM-Info*“, mit der die wissenschaftliche Grundlage für die Medikamentenänderung beschrieben wurde. Pro Patient und pro Brief wurde nur eine „*EbM-Info*“ eingefügt. Falls mehrere „*EbM-Infos*“ zutrafen, wählten die Assistenzärzte die Zusammenfassung mit der größten Relevanz für den Patienten. In der Kontrollgruppe wurde die

jeweils zutreffende „*EbM-Info*“ mit der Identifikationsnummer chiffriert. Jeder versandte Arztbrief wurde als Kopie archiviert.

5.2.2.3.5 Endpunkte und Endpunkterhebung

Als primären Endpunkt wählten wir die Absetzrate der Entlassungsmedikation bezogen auf die Praxen und adjustiert für diagnostische Gruppen. Wegen der postulierten Abhängigkeit zwischen diagnostischer Gruppe (z.B. Lipidstoffwechsel) und Absetzverhalten (z.B. Budgetgründe) hatten wir a priori festgelegt, dass für den primären Endpunkt eine Adjustierung erfolgen muss.

Sekundäre Endpunkte waren a) die Absetzrate nach Gründen („andere Meinung über die geeignete Behandlung; Veränderungen im Gesundheitszustand des Patienten, Patientenwunsch, Kostengründe, sonstiges“), bezogen auf die Arztbriefe und b) die Akzeptanz der „*EbM-Infos*“.

Endpunkterhebung: Nach mindestens drei Monaten Nachbeobachtung (in diesem Zeitraum wurde der Arzt bei einer N3-Packungsgröße mit ca. 100 Tabletten mindestens einmal vor die Entscheidung gestellt, das Medikament weiterzuführen oder abzusetzen) erfolgte die Endpunkterhebung in einem standardisierten Telefoninterview mit geschulten Interviewern. Der Fragebogen bestand aus einer Bewertung der „*EbM-Info*“ (bzw. der regulären Therapieempfehlung in der Kontrollgruppe) und soziodemographischen Faktoren. In der Interventionsgruppe wurden sieben Fragen, in der Kontrollgruppe vier Fragen auf einer 4-Item-Skala gemessen. Beide Gruppen erhielten zusätzlich acht identische Fragen über Kenntnisstand und Soziodemographie (s. Anhang 5.1). Inhalt, Verständnis, Eindeutigkeit der Fragen wurden mit mehreren unbeteiligten Ärzten in einem Telefoninterview vorgetestet und modifiziert. Das Interview dauerte für einen Patienten in der Interventionsgruppe fünf Minuten, in der Kontrollgruppe drei Minuten.

Vor der Analyse wurde von zwei verblindeten Ärzten überprüft, inwieweit die Aussagen der Hausärzte über Fortsetzung der Medikation und die angegebene laufende Medikation konsistent waren. Aufgrund einiger diskrepanter Angaben mussten wir weitere, vorab nicht geplante Analysen durchführen, die sich auch auf die Ergebnisse auswirkten.

5.2.2.3.6 Analyse und Powerberechnung

Primäre Analyse: Für die primäre Analyse wählten wir die Praxen (Cluster) als Sampling- und Evaluationseinheit. Bei der Randomisierung von Clustern besteht immer die Möglichkeit, dass sich Cluster von unterschiedlicher Größe unbalanciert auf die Gruppen verteilen^{298; 299}. Dies war auch in unserer Studie der Fall. Solange dadurch kein systematischer Fehler (z.B. unterschiedliche Teilnehmeraten) eingeführt wird, führen diese Unbalanciertheiten nicht zu einem Bias im Studienergebnis. Vor der Hauptanalyse wurden deshalb die Randomisierungsgruppen auf Vergleichbarkeit in den Ausgangsbedingungen getestet. Die Daten wurden mit Permutationstests analysiert. Bei Permutationstests müssen bestimmte Annahmen konventioneller statistischer Tests wie z.B. Normalverteilung nicht erfüllt sein. Die Tests untersuchen im Wesentlichen Symmetrieeigenschaften und eignen sich daher besonders für die Analyse clusterrandomisierter Studien ohne spezielle Verteilungsannahmen.

Die Sekundäranalysen über die Motivation für das Absetzverhalten wurden auf der Grundlage der Arztbriefe durchgeführt. Bei der Berechnung des Rater Agreement wurde Cohen's Kappa angewandt. Die Analyse wurde mit APL + Win 2.0 durchgeführt.

Power: Die Studie hatte eine 80% Power bei zweiseitigem alpha-Fehler von 0.05, um eine 20%ige Senkung der Absetzrate zu entdecken. Folgende Annahmen lagen der Berechnung der Studiengröße zugrunde: Geschätzte Absetzrate der Entlassungsmedikation in der Kontrollgruppe: 50%; klinisch bedeutsame Senkung der Absetzrate: 20% (von 50% auf 40%); verschiedene Annahmen über die Verteilungen von Ärzten / Briefen / Cluster (Praxis) auf der Grundlage der Zuweisungsdaten von 1999.

Ethische Aspekte: Die Hausärzte wurden vorab schriftlich und ein weiteres Mal zum Zeitpunkt des Telefoninterviews über die Studie informiert. Die Studie war von der Ethikkommission der Charité genehmigt worden.

5.2.2.4 Qualität der Arztbriefe - ein Audit

Die Kommunikation mit den einweisenden Hausärzten hat für die Innere Abteilung der PKW einen hohen Stellenwert. So wird z.B. großer Wert darauf gelegt, dass der Patient den Arztbrief zum Zeitpunkt der Entlassung erhält. Dazu wurde die Briefschreibung umgestellt, stark standardisiert und entsprechend elektronisch unterstützt. Eine Evaluation (Audit) über die Prozess- und Ergebnisqualität war bisher noch nicht durchgeführt worden. Da während der Planungsphase der „EbM-Infos“-Studie „Unzufriedenheit“ der niedergelassenen Ärzte mit den Arztbriefen der Inneren Abteilung als potentieller Confounder identifiziert worden war, wurde die Durchführung eines Audits beschlossen, in dem die einweisenden Ärzte eine Bewertung der Arztbriefe abgeben sollten³⁰⁰.

Zielsetzung war: a) die interne Evaluation des Ablaufs der Arztbriefschreibung für die PKW; b) Erfassung der für Hausärzte relevanten Elemente in einem Arztbrief; c) Ermittlung der Zufriedenheit mit den Arztbriefen. An alle Ärzte, die im Studienzeitraum in die Innere Abteilung der Parkklinik Patienten einwiesen, wurde einmalig ein standardisierter, voll anonymisierter Fragebogen mit Rückporto geschickt. Der Fragebogen war auf der Grundlage der (eher spärlichen) Literatur über Arztbriefschreibung^{293; 301-303} entwickelt worden und umfasste acht Fragen aus folgenden Bereichen: 1. Zustellungsdauer und Rückfrage, 2. Inhalte und Struktur (davon eine Frage mit 18 Optionen zur Wichtigkeit der einzelnen Items), 3. Umgang mit Klinikempfehlungen, sowie soziodemographische Informationen. Die Antworten für die einzelnen Items wurden auf einer Viererskalierung wiedergegeben. Der Fragebogen wurde an nichtbeteiligten Ärzten vorgetestet. Jeder einweisende Arzt erhielt nur einmal einen Fragebogen. Eine gemeinsame Analyse der Auditdaten und der Studiendaten wurde aus verschiedenen Gründen (u.a. Datenschutz) nicht vorgenommen.

5.2.2.5 Ergebnisse

5.2.2.5.1 Vergleichbarkeit der Gruppen

Die Charakteristika der teilnehmenden Ärzte, Praxen und Briefe sind in Tab. 5.1 und Abb. 5.4 zusammengefasst. Insgesamt konnten wir 178 randomisierte Praxen einschließen, 122 Praxen nahmen

am Interview teil. Damit wurde bei 268 / 417 Arztbriefen und über 59 / 135 unterschiedlichen „EbM-Infos“ ein erfolgreiches Interview durchgeführt (64%).

Tab. 5.1: Charakteristika der teilnehmenden Ärzte (Interviewebene)

Soziodemographie		EbM-Gruppe (n = 74 Praxen)	Kontrollgruppe (n = 58 Praxen)
Weiblich	68%	64%	73%
Alter (Jahre, Median / Range)	52 (34 - 66)	50 (36 - 65)	52 (34 - 66)
Hausärzte	68%	68%	68%
Allgemeininternisten	20%	18%	21%
Spezialisten	12%	13%	11%
Jahre praktizierend (Median / Range)	10 (0 - 40)	10 (1 - 14)	10 (0 - 40)
Gemeinschaftspraxis	42%	51%	31%

Vergleichbarkeit (s. auch Abb. 5.4): Bei der Durchführung der Studie wurden in der Interventionsgruppe mehr Entlassungsbriefe mit „EbM-Infos“ verschickt ($p=0.17$) und mehr Interviews durchgeführt ($p=0.03$). Eine Analyse auf selektive Beteiligung von EbM-Ärzten (Bias) an der Studie zeigte, dass die Unterschiede in der Teilnehmerate ausschließlich auf patientenbezogene Gründe (Versterben der Patienten) oder externe Gründe (Ungleichverteilung von vier Großpraxen mit > 10 eingeschlossenen Patienten: drei in der EbM-Gruppe und eine in der Kontrollgruppe) zurückzuführen waren. Arztbezogene Faktoren wie z.B. die Bereitschaft der Ärzte zum Interview waren zwischen den Randomgruppen nicht unterschiedlich verteilt.

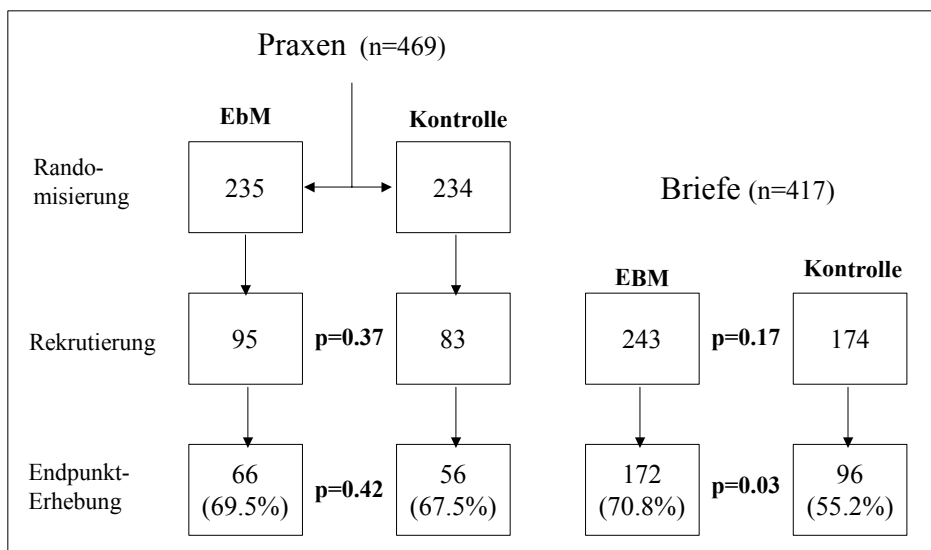


Abb. 5.4: Ergebnis von Randomisierung, Rekrutierung und Endpunkterhebung bezogen auf Praxen und Arztbriefe

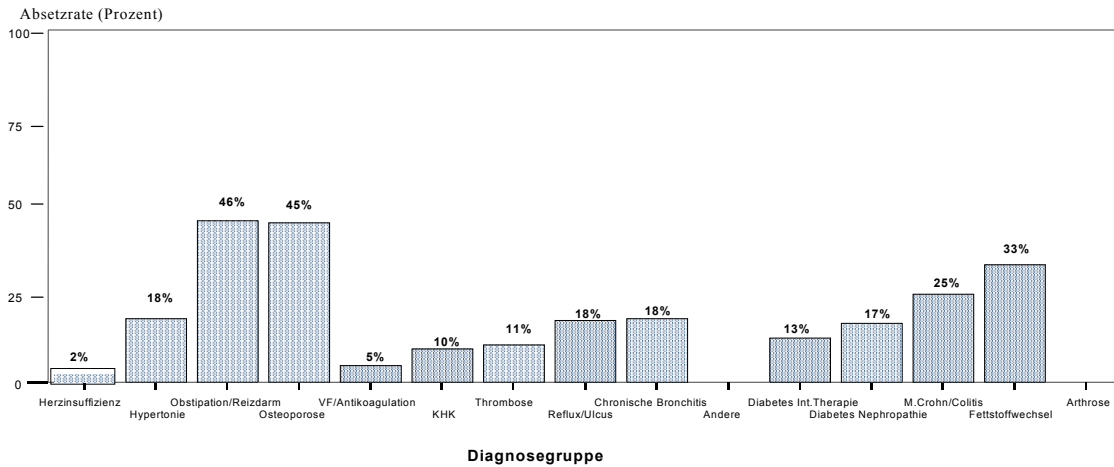
5.2.2.5.2 Einfluss der „EbM-Infos“ auf das Absetzverhalten der Hausärzte

„EbM-Infos“ und diagnostische Gruppen: In den Interviews wurden 59 „EbM-Infos“ in 15 diagnostischen Gruppen bewertet. Die „EbM-Infos“ waren bis auf drei (vermutlich zufallsbedingten) Ausnahmen gleichmäßig über die beiden Gruppen verteilt: Empfehlungen zu Herzinsuffizienz (13% vs. 21%) und Hypertonie (13% vs. 22%) befanden sich häufiger in der Kontrollgruppe, zu Osteoporose häufiger in der Gruppe mit den „EbM-Infos“ (10% vs. 6%).

Absetzrate der Entlassungsmedikation (Tab. 5.2)

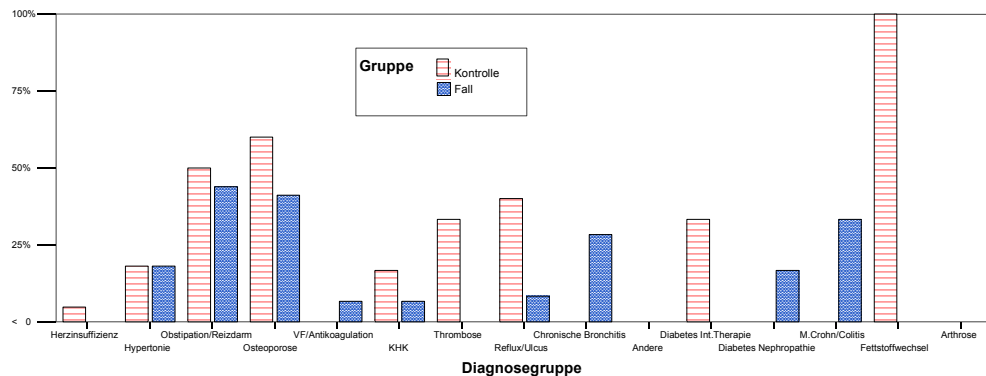
Ein Vergleich der Absetzraten zwischen den einzelnen Diagnosegruppen (Abb.5.5) bestätigt die in der Planungsphase geäußerte Vermutung einer Abhängigkeit zwischen Diagnosegruppe und Absetzverhalten. So schwankten die Absetzraten zwischen 0% (Arthrose) und 46% (Obstipation/Reizdarm).

Abb.5.5: Absetzraten nach Diagnosegruppe



Die Abbildung 5.6 zeigt das Absetzverhalten aufgeschlüsselt nach Diagnosegruppe und Exposition („EbM-Info“). Auch hier finden sich zwischen den einzelnen Diagnosegruppen erhebliche Schwankungen, die in der Hauptanalyse eine Adjustierung erfordern.

Abb.5.6: Absetzrate nach Diagnosegruppen und Exposition (EbM-Info vs. Kontrolle)



Die durchschnittliche Absetzrate / Praxis lag in der Kontrollgruppe bei 29,4% (d.h. deutlich unter der antizipierten Rate von 50%), in der „EbM-Info“- Gruppe bei 18,5%: In der „EbM-Info“-Gruppe war die adjustierte absolute Absetzrate („absolute Risikoreduktion“) der Medikamente um 12,5% ($p=0.039$) niedriger als in der Kontrollgruppe, der Effekt der Intervention größer als die in der Studienplanung antizipierte absolute Reduktion um 10%.

Tab. 5.2: Primärer Endpunkt: Absetzrate der Entlassungsmedikation

	EbM	Kontrolle	Absoluter Unterschied*	p-Wert
Absetzverhalten der Praxen	18.5%	29.6%	-12.5%	0.039

* adjustiert für diagnostische Gruppe

Die Qualitätskontrolle zu den Interviews ergab einige nicht weiter klärbare Diskrepanzen zwischen Entlassungsmedikation und laufender Medikation. Eine „worst-case“-Analyse auf der Grundlage der Plausibilitätsdaten nivellierte jedoch den Nutzen der „EbM-Infos“ in den nichtsignifikanten Bereich.

Motivation für das Absetzen der Medikamente (Tab. 5.3): Von den fünf vorgegebenen Gründe für ein Absetzen der empfohlenen Medikamente (andere Behandlungsvorstellungen des Hausarztes, patientenbedingte medizinische Gründe wie Verschlechterung des Zustands oder Nebenwirkungen, nichtmedizinische Gründe wie schlechte Compliance, zu hohe Medikamentenkosten oder Sonstiges) hatte nur ein einziger Grund einen signifikanten Einfluss auf eine Medikamentenänderung: „EbM-Infos“ führten dazu, dass teure Medikamente, die das Medikamentenbudget einer Praxis belasten würden, signifikant seltener abgesetzt wurden. (adjustierter absoluter Unterschied 5,8%; $p=0.01$).

Tab 5.3: Motivation für das Absetzen der Medikamente

	Absetzrate		Absolute Differenz*	p-Wert
	„EbM-Infos“	Kontrolle		
Andere Therapievorstellungen	2,3%	2,1%	-0.07%	0.97
Medizinische Gründe	10,5%	10,4%	-1,2%	0.74
Nicht medizinische Gründe	7,0%	6,3%	1,1%	0.71
Budgetgründe	1,7%	7,3%	-5,8%	0.014
Sonstige Gründe	1,2%	1,0%	-0,3%	0.89

* adjustiert für diagnostische Gruppe

5.2.2.5.3 Zufriedenheit der Hausärzte

Obwohl die „EbM-Infos“ für die Hausärzte in der Regel keine spektakulären Informationen über medizinische Neuentwicklungen enthielten, fiel die Bewertung der zusätzlichen Informationen durchweg positiv aus (Tab. 5.4). Insbesondere halfen die „EbM-Infos“ den Ärzten, die Gründe für die Verordnung besser zu verstehen, und sie erinnerten die Ärzte an eigentlich bekannte Therapieoptionen (z.B. Lipidsenkung bei Risikopatienten, Aspirin nach Apoplex, ACE-Hemmer bei Herzinsuffizienz) für diesen speziellen Patienten. Die meisten fühlten sich in ihrer Behandlungsentscheidung bestätigt und wünschten sich eine Fortsetzung der „EbM-Infos“.

Tab. 5.4: Bewertung der „EbM-Infos“ durch die Ärzte. Alle Fragen wurden auf einer 5-Item-Skala gemessen. Item 2 und 3 sowie Item 4 und 5 wurden in dieser Tabelle gemeinsam ausgewertet.

„EbM-Infos“ ...	Nein	Etwas	Stark / sehr stark
... enthielten neue Informationen?	80%	10%	10%
... halfen, die Krankenhausempfehlungen besser zu verstehen?	17%	65%	18%
... hatten einen Einfluss auf die Fortsetzung der Behandlung?	46%	42%	12%
... waren ein Eingriff in die ärztliche Freiheit?	85%	12%	3%
... sollten fortgesetzt werden?	2%	26%	72%

5.2.2.5.4 Audit: Zufriedenheit der einweisenden Hausärzte mit der Arztbriefqualität

Parallel zur Rekrutierung der Patienten für die randomisierte Studie (1.5.– 31.12.2000) wurden an alle einweisenden Ärzte 362 Fragebögen verschickt. 64% der Einweiser waren Ärztinnen, 36% Ärzte, was in etwa der Geschlechterverteilung niedergelassener Ärzte in Ostdeutschland entspricht.

Der Rücklauf der Umfrage lag bei 44% (159/362 Fragebögen); 83% kamen von Hausärzten (Internisten und Allgemeinmediziner), 14% von Fachärzten (inklusive der als Facharzt tätigen Internisten). Die allgemeine Zufriedenheit mit dem Arztbrief war sehr groß: 97% äußerten sich als zufrieden bis sehr zufrieden. Die Ärzte gaben an, dass 77% der Briefe von den Patienten mitgebracht wurden, 14% trafen innerhalb von einer Woche ein, so dass für 91% der Patienten eine zeitnahe Informationsübermittlung stattfand. Von den 18 Optionen der für den Hausarzt essentiellen Informationen in einem Entlassungsbrief wurden am häufigsten der Aufklärungsgrad des Patienten, die empfohlene Medikation und die Mitteilungen von erhobenen Befunden genannt. Auf die Frage nach dem Umgang mit Klinikempfehlungen zur Weiterbehandlung gaben 80% an, dass sie die Empfehlungen in 50% und mehr Fällen übernehmen würden. Bei den Gründen für die Änderung (Mehrfachnennung möglich) wurde am häufigsten Unwirtschaftlichkeit (72%) genannt, gefolgt von zu großer Komplexität der Therapieempfehlungen (30%). Abweichende fachliche Meinung (15%) oder Patientenwunsch (13%) rangierte am unteren Ende.

Den Angaben ist nicht zu entnehmen, wie mit Änderungen wegen Unwirtschaftlichkeit umgegangen wurde, ob z.B. eine Umstellung auf ein Generika erfolgte, teure Medikamente ersatzlos abgesetzt oder andere als gleichwertig erachtete, aber kostengünstigere Behandlungsstrategien gewählt wurden (z.B. H2-Blocker statt Protoneninhibitor). Allerdings zeigt die Umfrage eindeutig, dass die Akzeptanz der Krankenhausempfehlungen unter den Hausärzten wesentlich größer ist als für die Berechnung der Fallzahl des RCTs angenommen.

5.2.2.6 Diskussion

5.2.2.6.1 Zugewinn der Studie

Die „EbM-Infos“ in Arztbriefen hatten in unserer Studie eine signifikante und klinisch relevante Auswirkung auf die Entscheidungsfindung niedergelassener Ärzte und waren gut akzeptierte Instrumente, um EbM in die tagtägliche Praxis einzuführen. Unsere Studie widerlegt die Ansicht, dass Ärzte die evidenzbasierte Medizin per se ablehnen²⁶⁶. Vielmehr bestätigt sie den Bedarf für pragmatische Lösungen^{216; 236; 304}, um qualitativ hochwertige Forschungsergebnisse in die Versorgung zu integrieren und die Praxis entsprechend anzupassen. Für viele in der Literatur geäußerte Bedenken gegenüber EbM wurde in unserer Studie Sorge getragen: Die Information bezog sich auf einen ganz spezifischen Patienten²⁷², war zum Zeitpunkt der Entscheidung verfügbar^{273; 304} und wurde zusätzlich durch die Autorität des Chefarztes aus dem lokalen Krankenhaus unterstützt, der den Patienten selbst kannte³⁰⁵. Das Zusammenspiel dieser Faktoren hat sich aus unserer Sicht positiv auf die Fortführung der angesetzten Medikamente ausgewirkt und wesentlich zum Erfolg der Intervention beigetragen.

5.2.2.6.2 Methodische Herausforderungen der Studie

Studien zu Interventionen im Gesundheitswesen können um ein Vielfaches komplexer und vielschichtiger werden als randomisierte Studien über medikamentöse Behandlungen, was in Design und Analyse berücksichtigt werden muss^{34; 296}. Auf einige methodische Besonderheiten und Herausforderungen, die bei der Durchführung unserer Studie auftraten, soll hier im Weiteren eingegangen werden:

Ungleichverteilung in den Clustern: Studien in organisatorischen Einheiten verlangen Clusterrandomisierung³⁵. Seit dem zunehmenden Interesse an randomisierten Evaluationen auch in der (Versorgungs-) Forschung wird diese in der Medizin relativ neue Randomisierungsmethode gehäuft eingesetzt, um Einheiten zu beobachten, bei denen durch Gemeinsamkeiten Gruppeneffekte und Abhängigkeiten entstehen. Allerdings ist dieses Zuteilungsverfahren auch mit Nachteilen verbunden: Trotz einer balancierten Randomisierung auf der vorab festgelegten Auswertungsebene (hier die Praxis) finden sich häufiger zufallsbedingte Ungleichverteilungen auf der Interventionsebene (hier der Arztbrief bzw. der Patient). Diese sind auch in der Literatur beschrieben und muten auf den ersten Blick wie Bias an^{298; 299}. Dies war der Fall in unserer Studie. Solange allerdings sichergestellt werden kann, dass keine systematische Selektion stattgefunden hat, reduzieren solche Ungleichheiten zwar die Power, führen aber nicht zu einer Verzerrung des Ergebnisses. Detaillierte Analysen über die Mechanismen und Hintergründe über die Nichtteilnahme an der Studie zeigten, dass die beobachteten Ungleichheiten zum einen durch eine diskrete Ungleichverteilung auf Clusterebene zustande kamen, aber im Wesentlichen auf patientenbezogene Gründe (wie Versterben) und externe Gründe (wie die zufallsbedingte Ungleichverteilung der drei von vier Großpraxen in die EbM-Gruppe) zurückzuführen sind. Es ist daher nicht anzunehmen, dass die beobachteten Interventionseffekte zuteilungsbedingt sind.

Berechnung von Power und Studiengröße: Für konfirmatorische Studien ist eine auf Powerüberlegungen beruhende Fallzahlplanung unerlässlich. Die dazu benötigten Zahlen stehen im Idealfall aus früher durchgeführten Studien zur Verfügung, im ungünstigsten Fall müssen sie aus Erfahrungen über den Daumen geschätzt werden. Die Annahmen unserer Studie über die Abbruchraten stammten zum Teil aus der Literatur²⁹⁵. Für die Schätzungen der Clustergröße wurde eine Analyse der Einweisungspraxen der Vorjahre mit herangezogen, die Rekrutierungsraten wurden auf der Grundlage der Einweisungszahlen und einer orientierenden Stichprobe von Entlassungsbriefen berechnet. Zum Teil waren die zugrundegelegten Annahmen jedoch über den Daumen gepeilte Schätzungen aufgrund fehlender Vorinformationen. Während der Durchführung der Studie zeigten sich mehrere Abweichungen: Die Rekrutierung verlief langsamer, und trotz Verlängerung wurde die ursprünglich geplante Praxenzahl nicht erreicht. Über 20% der geeigneten Patienten wurden bei der

Routineversorgung versehentlich nicht eingeschlossen (allerdings ohne Hinweis auf Selektionsbias, wie die nachträglich durchgeführte Überprüfung ergab). Die Größe der Cluster variierte stärker als ursprünglich angenommen. Die größte Beeinträchtigung der Power stammte jedoch von der deutlich besseren Akzeptanz der Krankenhausempfehlungen in der Kontrollgruppe (70% statt der erwarteten 50%). Diese ungünstige Auswirkung auf die Power der Studie lässt sich am besten mit folgenden Überlegungen veranschaulichen: Wenn man eine Validierungsstudie auf der Grundlage der Ergebnisse unserer Untersuchung planen würde, so müsste diese Studie doppelt so groß angelegt werden wie unsere ursprünglich geplante Studie oder achtmal so groß wie die tatsächliche durchgeführte Studie. Da jedoch unsere Primäranalyse (trotz geringerer Power) ein statistisch signifikantes und klinisch bedeutsames Ergebnis zeigte, sind diese Überlegungen weniger für die Interpretation dieser Studie als vielmehr für die Planung ähnlicher zukünftiger Studien relevant.

5.2.2.6.3 Weitere Betrachtungen

Das Gesamtprojekt gehört mit seiner Zielsetzung zu den „unspezifischen“ EbM-Projekten, die unter 4.1.1.1 beschrieben wurden. Es wurde eine Infrastruktur für evidenzbasierte Praxis aufgebaut, Ärzte setzten sich in Schulungen und während der Erstellung der „*EbM-Infos*“ mit Methodik und Theorie der EbM auseinander und praktizierten in den Diskussionen auf den Visiten und mit der Anwendung der „*EbM-Infos*“ in den Arztbriefen eine konkrete Form, EbM in die tägliche Versorgung umzusetzen. Wir vermuten, dass das intensive und praxisnahe Projekt zu einem Lerneffekt und zu Änderungen im Verhalten der Ärzte geführt hat. Inwieweit diese Annahmen stimmen, kann wegen der fehlenden systematischen Begleitevaluation nicht beantwortet werden. In dem Abschnitt „Implementierung in Deutschland“ wird im Zusammenhang mit anderen Klinikprojekten auf die Übertragbarkeit der Ergebnisse und die Konsequenzen für weitere Projekte eingegangen.

5.3 EbM-Informationsservice: Ein Pilotprojekt mit Rechercheangebot und kritischer Literaturbewertung (eigene Untersuchungen)

Eine andere Form, eine „evidenzbasierte“ Patientenversorgung zu unterstützen, wurde in einem Pilotprojekt am Institut für Sozialmedizin der Charité (Projektleitung Dr. U. Euler) erprobt: „Der EbM-Informationsservice“. Dabei erhielten Ärzte, die ihre Patienten nach EbM-Kriterien behandeln wollten, ohne aber selbst die Literaturrecherche und die kritische Bewertung durchführen zu müssen, das Angebot, durch einen externen Informationsservice auf spezifische Patientenprobleme „evidenzbasierte Antworten“ zu bekommen.

5.3.1 Fragestellung

Ziel des Projekts war, die Machbarkeit und Akzeptanz eines externen Rechercheangebots zu patientenbezogenen Fragen mit kritischer Studienbewertung an einem Universitätsklinikum zu überprüfen. Zusätzlich wollten wir Informationen über die Nutzer des Services, die gestellten Fragen und die Konsequenzen für die Entscheidungsfindung erheben.

5.3.2 Methodik

Die Durchführung des Projekts erfolgte in zwei Phasen: 1. Recherche, kritische Literaturbewertung und strukturierte Beantwortung der Fragen; 2. Erhebung von Zufriedenheit und Nutzen der Ärzte durch einen standardisierten Fragebogen.

Recherchephase: Über einen zentralen Emailverteiler wurden alle ärztlichen Mitarbeiter der Charité einmalig über einen sechswöchigen Rechercheservice informiert. Der Service beinhaltete das Angebot, ein formuliertes klinisches Problem zeitnah nach EbM-Kriterien zu beantworten. Das Problem sollte als strukturierte Frage mit den vier Komponenten 1. Patient/Problem 2. Intervention / Exposition 3. Vergleichsintervention /-exposition und 4. Zielgröße vorgelegt werden. Die eingehenden Fragen wurden kategorisiert nach: Therapie/Prävention, Ätiologie, Diagnostik/Screening, Prognose und Nebenwirkungen. Zwei Fachärzte (U. Euler, A. Juche) mit umfangreicher Erfahrung in EbM und Recherchekompetenzen führten in fünf Datenbanken mit zusammengefasster bzw. vorbewerteter

Evidenz die Suche durch. Bei unbeantworteten Fragen wurde die Suche auf Medline ausgedehnt. Die Studie mit der größten klinischen Relevanz und der höchsten Evidenzstufe wurde nach der EbM-Vorgehensweise kritisch bewertet¹⁹⁵. Die Fragen wurden nach folgender Struktur beantwortet: 1. Fragestellung 2. reproduzierbare Suchstrategie 3. kritische Bewertung mit quantitativem Ergebnis (falls möglich) und 4. Zusammenfassung und Angabe der Evidenzstufe.

Evaluationsphase: Ein bis zwei Wochen nach der Beantwortung wurde mit einem standardisierten Fragebogen (s. Anhang 5.2) die Bewertung abgefragt: Zufriedenheit und Nutzen der Information, Transparenz und Verständlichkeit der Antwort, Einfluss auf die Entscheidungsfindung sowie teilnehmerspezifische Angaben (Berufserfahrung, EbM-Erfahrungen). Die Ergebnisse wurde deskriptiv analysiert.

5.3.3 Ergebnisse

Von den 34 eingegangenen Anfragen waren 31 als spezifische, problemorientierte Frage strukturiert. Der Rücklauf der Evaluationsfragebögen lag bei 61% (19/31).

Recherche: Am häufigsten wurde der EbM-Literaturservice von Ärzten in leitenden Funktionen (Alter: 41 Jahre; Dienstjahre (Median): 14 Jahre) in Anspruch genommen. Tabelle 5.5 beschreibt die Fachrichtungen der Teilnehmer und die Zuordnung der Fragen zu klinischen Themenbereichen (Tab.5.5).

Tab.5.5: Fachrichtung der Teilnehmer und Themenbereiche der gestellten Fragen

Fachrichtung der teilnehmenden Ärzte (n=31)	Klinisches Problem
nichtoperativer Bereich: 43%	Therapie: 77%
operativer Bereich: 23%	Ätiologie: 10%
Zahnmedizin: 20%	Prognose: 6%
andere Bereiche: 13%	Diagnose: 3%
	Nebenwirkungen: 3%

Die durchschnittliche Bearbeitungsdauer einer Frage betrug sieben Stunden (Median, Range: 3-32,5h), wobei 40% der Zeit auf die Literatursuche, 40% auf die Bewertung und 20% auf die Beantwortung der

Frage fielen. Bei Einordnung der Studien nach Evidenzstufe³⁰⁶ zeigten sich für eine Frage eine Ia-Antwort und für zwei Fragen eine Ib-Antwort. Je sieben Fragen wurden mit IIb bzw. IV-Evidenz beantwortet, für zwei Fragen fand sich Level V-Evidenz. Bei zwölf Fragen war eine weitere Zuordnung der Evidenz nicht möglich.

Evaluation: Das Rechercheangebot wurde sehr positiv aufgenommen. Für zwei Drittel der Fragesteller (13/19) waren die Antworten hilfreich, transparent und entsprachen dem benötigten Informationsbedarf, dennoch hatten die Therapieempfehlungen nur moderaten bis keinen Einfluss auf die Entscheidungsfindung.

Die wichtigsten Quellen für neue Informationen waren Fachzeitschriften (100%), Fortbildungen (58%) und Kongresse (58%). Von den elektronischen Datenbanken wurden primär Medline (95%) und das Internet eingesetzt. 26% benutzten die Cochrane Library, 16% die Literaturdienste bei DIMDI und eine Person Clinical Evidence (5%). 4/19 Teilnehmer hatten bereits an einem EbM-Kurs teilgenommen.

5.3.4 Schlussfolgerungen

Ein Service für klinische Fragestellungen ist an einem Universitätsklinikum durchführbar und wird in der angebotenen Form als hilfreich, transparent und didaktisch wertvoll empfunden. Allerdings deuten die Teilnehmercharakteristika (leitende Ärzte und EbM-Interessierte) und die z.T. sehr speziellen Anfragen auf eine ausgesprochene Selbstselektion hin. Die Beantwortung der Fragen einschließlich der detaillierten Dokumentation war sehr aufwändig. Angesichts des Aufwands sollte man sich auf häufiger auftretende Probleme beschränken und die recherchierten Antworten in einer lokalen Datenbank auch für anderen Ärzten zugänglich machen. International halten inzwischen viele Kliniken oder Abteilungen die Ergebnis ihrer Journal Club-Besprechungen als CAT (Critically Appraised Topics)-Datenbanken im Internet verfügbar (<http://cebm.jr2.ox.ac.uk/> oder www.google.de „Critically appraised topics“ „CAT“).

Das Projekt wurde an einer Universitätsklinik mit großem Fundus an Online-Literatur durchgeführt. An nichtuniversitären Kliniken oder im niedergelassenen Bereich mit beschränktem Zugriff auf die Primärliteratur kommt einem entsprechenden Informationsservice oder der Verfügbarkeit von

vorbewerteter Evidenz einschließlich (evidenzbasierten) Leitlinien eine wesentlich größere Bedeutung zu. Das bestätigen zwei Projekte aus Großbritannien^{304; 307}, in denen Allgemeinärzten einer definierten Region ein schneller, pragmatisch orientierter Informationsservice angeboten wurde. Quintessenz beider Studien war, dass evidenzbasierte Informationen von den Hausärzten als Entscheidungsgrundlage sehr geschätzt wurden, sie sich aber Zugang zu den Information wünschten, ohne auf eigene Fertigkeiten in Literatursuche und -bewertung angewiesen zu sein.

5.4 Implementierung in Deutschland

5.4.1 Praxisprojekte

Im Gegensatz zu den evidenzbasierten Disease-Management-Programmen, wie sie derzeit von den Krankenkassen im Gesundheitswesen entwickelt und eingeführt werden, entstanden das PKW-Projekt und ähnlich gelagerte Projekte (St. Antonius-Krankenhaus, Wuppertal Tab. 5.6¹¹⁴, St. Franziskus-Hospital, Köln) aus Initiativen der beteiligten Ärzte. Alle Ärzte, vom Arzt im Praktikum bis zum Chefarzt, waren an der Entwicklung und Umsetzung der Projekte im Sinne einer „Ownership“ beteiligt. In dem Projekt des St. Antonius-Krankenhaus wie in unserem Projekt wurden sektorenübergreifend auch die niedergelassenen Ärzte miteinbezogen. In diesem Sinne entstand entsprechend der Wurzeln von EbM ein Projekt „von Ärzten für Ärzte“¹⁰⁰ und wir konnten zeigen, dass sich EbM im ärztlichen Alltag trotz knappem Zeitrahmen praktizieren lässt. Allerdings sind strukturelle Voraussetzungen unabdingbar. Durch Anbindung an das gut funktionierende Qualitätsmanagement und eine seit längerem bestehende Qualitätsdiskussion in der PKW waren Ärzte und Personal für die Thematik sensibilisiert, was vermutlich wesentlich zur Umsetzung des Projekts beigetragen hat. Sicher lassen sich die Ergebnisse nur bedingt auf Kliniken mit weniger weit entwickelter Infrastruktur und Problembewusstsein übertragen. Mit der Integration der „EbM-Infos“ in die Arztbriefe wurde eine explizite „evidenzbasierte“ Medizin an der Schnittstelle stationär / ambulant angesiedelt und mit Erfolg (d.h. mit Einfluss auf das Verhalten der Ärzte) in die ambulante Versorgung weitergetragen. Den unmittelbaren Lerneffekt, den die niedergelassenen Ärzte bei den

präzisen patientenspezifischen Informationen empfanden, sehen wir als wichtigen Faktor für die gute Akzeptanz.

Tab. 5.6: Elemente des Wuppertaler EbM-Projekts¹¹⁴

<ul style="list-style-type: none"> • hausinterner EbM-Grundkurs • Tutorengruppen mit Bildung „klinischer Tandems“, • Journal Club • tägliche Kurzvorträge • Integration anderer Fachabteilungen und Berufsgruppen in die EbM-Aktivitäten • neue Formen von Fortbildungsveranstaltungen an der Schnittstelle zur ambulanten Medizin • Planung und Durchführung weiterer Wuppertaler EbM-Grundkurse 	<ul style="list-style-type: none"> • Aufbau einer Infrastruktur (Computernetzwerk und Bereitstellung von EbM-Ressourcen) • Erstellung einer sektorenübergreifenden elektronischen Dokumentationsplattform mit EbM-gestützten Leitlinien (Q-med) • Entwicklung / Einführung klinikinterner Leitlinien • Wissenschaftliche Projekte
<p>Evaluierung: Zum Zeitpunkt der Projektbeschreibung (2000) war keine Evaluierung geplant; Option einer nachträglichen Evaluierung vorbehalten</p>	

5.4.2 Voraussetzungen für eine erfolgreiche Implementierung.

Die evidenzbasierte Medizin als Brückendisziplin zwischen der patientenorientierten Forschung und der Patientenversorgung erfordert Expertise in klinischen Fächern, ein Grundverständnis in klinischer Epidemiologie sowie Kenntnisse im Erwachsenenlernen (Kapitel 4). So ist es nicht überraschend, dass sich im internationalen Vergleich viele EbM-Zentren an Institute für klinische Epidemiologie oder Institute für Versorgungsforschung („*Health Services Research*“) mit multidisziplinärer Ausrichtung angebunden haben. Um die Praxisrelevanz und die Bedeutung des Fachs für die Versorgung zu signalisieren, sollten Vertreter und Meinungsführer der klinischen Fachbereiche aktiv eingebunden und integriert werden.

Wird eine weitere Verbreitung der genuinen EbM-Arbeitsweisen in der Versorgung angestrebt, müssen die Grundkonzepte stärker in die Ärzteschaft hineingetragen, vorhandene Missverständnisse ausgeräumt und der Nutzen und Gewinn für den einzelnen Arzt besser vermittelt werden.

Folgt man der Literatur, müssen dazu drei Voraussetzungen erfüllt werden:

a) Schulung und Ausbildung (s. auch Kapitel 4) aller Gesundheitsberufe mit geeignetem Lehrmaterial, Programmen und Modulen über die Kerninhalte der EbM, angepasst an den jeweiligen Ausbildungsstand und an die Bedürfnisse. Dies ist ein mehrstufiger Prozess, der bereits in der Ausbildung beginnen und als fester Bestandteil in die Weiter- und Fortbildung aller Gesundheitsberufe integriert werden sollte. Es gibt inzwischen ausreichend Literatur, dass viele Ärzte die Primärliteratur nicht selbstständig suchen und kritisch bewerten möchten. Ziel dabei sollte nicht primär die Ausbildung von Experten mit vertieften EbM-Fertigkeiten sein, sondern von Ärzten mit einem kritischen Verständnis von EbM und den Vorteilen vorbewerteter Evidenz und mit einigen Grundfertigkeiten wie dem Formulieren strukturierter Fragen oder dem Auffinden vorbewerteter Evidenz.

b) Zugriffsmöglichkeiten auf EbM-Ressourcen („Pre-appraised evidence“), bei denen der zeitaufwändige Bewertungsprozess von Methodikern bereits vorgenommen wurde. Bei der in der Regel mäßig guten Ausstattung nichtuniversitärer Krankenhäuser mit Primärliteratur (oft verfügen sie nicht einmal über eine allgemein zugängliche Bibliothek) ist der Zugriff auf aktuelle Datenbanken mit vorbewerteter Evidenz einschließlich evidenzbasierter Leitlinien umso wichtiger. Gerade effiziente schnelle und hochwertige Informationsmöglichkeiten am „Point of Care“ haben sich für eine breite Akzeptanz von EbM und eine rasche Integration in die Routineversorgung als essentiell erwiesen. Im Internet existiert bereits eine Vielzahl von themenspezifischen, vielfach kostenlos zugänglichen Datenbanken (www.ebm-netzwerk.de). Die aufwändige Aufarbeitung, Bereitstellung und regelmäßige Aktualisierung solcher Ressourcen lassen sich allerdings nur im Rahmen von größeren Kooperationen (vermutlich auf internationalem Niveau) erzielen. Beispielhaft ist in diesem Zusammenhang auch die eLibrary des NHS www.nelh.nhs.uk.

c) Hilfen bei der Umsetzung: Wissenschaftliche Informationen können auf vielerlei Art systematisch in die Patientenbehandlung übernommen werden. Nimmt man die Ergebnisse der beschriebenen Studien ernst, muss allerdings wesentlich mehr Aufmerksamkeit auf eine aktive Umsetzung in die Praxis und die lokalen Möglichkeiten und Bedürfnisse gelegt werden^{114; 115}: Dazu gehören Barrierenanalysen am Arbeitsplatz, Bereitstellung der Infrastruktur und Ressourcen, Entwicklung von

Anwendungsmodulen zur Umstellung der eigenen Arbeitsweise oder der ganzen Teams. Die Verbindungen zum Qualitätsmanagement sind offensichtlich und die Zusammenarbeit müsste konsequent weiterentwickelt und ausgebaut werden. Lokale Initiativen brauchen Expertenunterstützung für die Planung und erfolgreiche Umsetzung von EbM in ihrer jeweiligen Arbeitsumgebung. Möglichkeiten für einen aktiven Erfahrungsaustausch unter den Ärzten, die in ihren Kliniken häufig Einzelkämpfer sind, müssen bereitgestellt werden. Für diese Anliegen hat sich im Deutschen Netzwerk EbM eine Arbeitsgruppe „EbM in Klinik und Praxis“ gebildet.

6 Zusammenfassung und Perspektive

Die vorliegende Habilitationsschrift behandelt das Thema „Die klinische Epidemiologie in der ärztlichen Entscheidungsfindung“. In dem **ersten Projekt** „Beobachtung oder Experiment“ haben wir den Nachweis geliefert, dass die Randomisierung für den Wirksamkeitsnachweis klinischer Interventionen unverzichtbar ist, um eine möglichst unverzerrte Effektmessung sicherzustellen. In dem **zweiten Projekt** haben wir in einer Simulationsstudie auf der Grundlage realer Patientendaten das Potential für klinische Fehlentscheidungen durch Bias demonstriert und die Risikokonstellationen identifiziert, die für diese Gefahr besonders anfällig sind: Moderates bis niedriges Patientenrisiko und moderate bis geringe Wirksamkeit der medizinischen Intervention. Wegen des häufigen Auftretens dieser Konstellationen in der Patientenversorgung haben die Erkenntnisse unserer Studie einen ganz konkreten Praxisbezug. In unserer Untersuchung zum Leitlinien-Clearingverfahren Hypertonie haben wir gezeigt, dass diese für die Versorgung von Patienten äußerst relevanten Erkenntnisse beim Formulieren von Praxisempfehlungen nicht ausreichend berücksichtigt werden. In dem **dritten Projekt**, der Berliner EbM-Studie, haben wir auf der Grundlage des von uns konzipierten Berliner Gegenstandskatalogs EbM ein Instrument entwickelt und validiert, mit dem wir zuverlässig und reproduzierbar zwischen unterschiedlichen Kenntnissen und Fertigkeiten von EbM differenzieren können. In einer dreijährigen Studie haben wir nachgewiesen, dass durch kurze intensive Kurse in evidenzbasierter Medizin (wie z.B. unsere Berliner EbM-Kurse) bei den Teilnehmern ein statistisch signifikanter und klinisch relevanter Wissenszuwachs erzielt wird (vor dem Kurs $6,3 \pm 2,9$, nach dem Kurs $9,9 \pm 2,4$ richtige Antworten bei insgesamt 15 Fragen; $p < 0.001$). In dem **vierten Projekt**, einer clusterrandomisierten Studie unter Hausärzten, hatten wir untersucht, ob kurze, evidenzbasierte Erläuterungen für im Krankenhaus neu angesetzte Behandlungen Hausärzte motivieren können, diese Behandlung fortzusetzen. Dabei fanden wir eine statistisch signifikante Reduktion der Absetzrate (absolute Risikoreduktion 12,5%; $p=0.039$).

Mit den einzelnen Projekten wurde ein weiter Bogen geschlagen, von den methodischen Grundlagen der klinischen Epidemiologie, über ihre Bedeutung für differenzierte ärztliche Entscheidungen und für

die Vermeidung offensichtlicher Fehlentscheidungen. Sowohl am Beispiel der Leitlinienentwicklung wie auch bei unserer Untersuchung zum Nutzen der Berliner EbM-Kurse fanden wir bei den teilnehmenden Ärzten bedeutende Defizite in EbM-Basiswissen, die allerdings durch strukturierte Fortbildungen wie den EbM-Kursen deutlich verbessert werden konnten. Für die Umsetzung von EbM in der Patientenversorgung gibt es bisher in Deutschland nahezu keine Untersuchungen. Der Einsatz von EbM-Infos in Arztbriefen in unserer clusterrandomisierten Studie als Teil des EbM-Projekts der Park-Klinik Weissensee war ein erster Versuch, den Praxiseinsatz von EbM zu untersuchen.

Im Folgenden führe ich die Erkenntnisse der einzelnen Studien in einer übergreifenden Diskussion zusammen:

Mit der klinischen Epidemiologie als Mutterdisziplin verfügt die evidenzbasierte Medizin über ein robustes methodisches Instrumentarium, mit dem sie die Zuverlässigkeit der Ergebnisse aus der patientenorientierten Forschung beurteilen kann. Die Bestätigung der logischen Vorteile der Randomisierung in unseren Untersuchungen hat unmittelbare Konsequenzen für die Praxis der EbM. Die Studien untermauern die Gültigkeit zentraler EbM-Konzepte, wie z.B. die Evidenzstufen („Levels of Evidence“) oder die darauf aufbauenden Empfehlungsstärken („Strength of Recommendations“). Dabei besagen die Evidenzstufen, dass sich die verschiedenen Studiendesigns in der Glaubwürdigkeit ihrer Ergebnisse unterscheiden und legitimieren diese Hierarchie mit den eingesetzten methodischen Techniken, die vor einer Verzerrung der Ergebnisse schützen.

In Gesundheitssystemen, die zunehmend höhere Ansprüche an die Evidenzbasis von Maßnahmen stellen, bevor sie diese als Leistungen anbieten, haben diese Hierarchien einen erheblichen Einfluss auf gesundheitspolitische Entscheidungen: In evidenzbasierten Leitlinien werden danach Maßnahmen und Interventionen eingestuft und empfohlen oder abgelehnt. In den Gremien der deutschen Selbstverwaltung finden diese Grundlagen Eingang in die Entscheidungsfindung, ob Maßnahmen in den Leistungskatalog aufgenommen oder abgelehnt werden. Mit der zunehmenden Verbindlichkeit der Disease-Management-Programme beeinflussen die Leitlinien die Behandlung Tausender von Patienten. Allerdings ist der Einsatz von vereinfachten Schemen der Evidenzklassifikation, die z.B. ausschließlich auf dem Zuteilungsverfahren der Patienten beruhen, zunehmend und berechtigterweise

in Kritik geraten, da andere wichtige Quellen der Effektverzerrung nicht berücksichtigt werden. Ein Vorschlag für eine differenzierte Klassifikation von Evidenzstufen wurde von uns kürzlich vorgestellt¹⁵⁹.

Diese methodischen Überlegungen sind nicht nur fundamental für Empfehlungen, wie sie in Leitlinien zur Qualitätssicherung oder in Disease-Management-Programmen zur Steuerung des Gesundheitssystems ausgesprochen werden. In unserer Simulationsstudie (2.2) haben wir entwickelt, wie Bias in Studien die Entscheidungsfindung der Ärzte bei individuellen Patienten unmittelbar beeinflussen kann, wobei das Potential für Fehlentscheidungen zusätzlich von dem Zusammenspiel von Patientenrisiko und Behandlungseffekt abhängt. Für die Versorgung im Allgemeinen hat diese Beobachtung ganz konkrete Auswirkungen, da die wirklich kritische Konstellation - Niedrigrisikopatient und mäßig wirksame Intervention - im Alltag häufig anzutreffen ist. Zusätzlich ist ein Trend zu beobachten, dass immer kleinere Risiken behandelt werden, selbst wenn der daraus resultierende Behandlungsnutzen nur noch marginal ist. Wie aus der Literaturübersicht hervorgeht, ist dieser Zusammenhang vielen behandelnden Ärzten nicht bewusst und wird trotz dringender Empfehlung von den wissenschaftlichen Fachgesellschaften nicht angemessen und mit dem nötigen Nachdruck unterstützt. Hier kann die EbM den Meinungsbildnern in der Medizin, den praktizierenden Ärzten und den Patienten wirkungsvoll Werkzeug und Orientierung bieten, um in spezifischen Versorgungssituationen den Behandlungsnutzen für den einzelnen Patienten zu erfassen. Bei dem gegenwärtigen Trend im Gesundheitswesen, die Patientenversorgung durch Leitlinien stärker zu standardisieren, gewinnen diese Fertigkeiten zunehmend an Bedeutung, da Leitlinien nur Empfehlungen für „Patientengruppen“ abgeben können, die der einzelne Arzt wieder individuell auf den Patienten anpassen muss. An diesen Beispielen lässt sich auch eindrucksvoll und ganz konkret demonstrieren, dass evidenzbasierte Medizin mehr bedeutet als ein inflationärer Einsatz von Studienzitaten.

Allerdings werden die für die Praxis einer evidenzbasierten Medizin erforderlichen Fertigkeiten in der Ärzteschaft nicht ausreichend beherrscht, wie wir in unserer Studie zum Nutzen von EbM-Kursen nachweisen konnten, in denen das EbM-Basiswissen von hochmotivierten Ärzte mit großem Interesse an EbM vor dem Kurs nur marginal besser war als das von Medizinstudenten in den unteren

Semestern, die mit EbM noch nicht in Berührung gekommen waren. In dieser Situation stellt sich die Frage, ob mit den Disease-Management-Programmen nicht doch eine „Kochbuch-Medizin“ etabliert wird, wenn nicht parallel dazu die EbM-Fertigkeiten des einzelnen Arztes auf- und ausgebaut werden. Die Forderung nach intensiven Schulungen der Ärzteschaft wird zusätzlich gerechtfertigt, da wir in unserer mehrjährigen Studie nachweisen konnten, wie sich der Kenntnisstand durch Schulungen wie den von uns veranstalteten Berliner EbM-Kursen erheblich verbessern lässt. Mit dem Curriculum Evidenzbasierte Medizin der Ärztlichen Zentralstelle Qualitätssicherung und des Deutschen Netzwerks EbM existiert bereits der Rahmen für ein standardisiertes Fortbildungsprogramm.

Die dokumentierten Wissenslücken sind sicher eine Erklärung für die zahlreichen Missverständnisse innerhalb der Ärzteschaft über die Inhalte der EbM und ihre Konsequenzen für die Versorgung. Es gibt wenig Anlass anzunehmen, dass sich eine evidenzbasierte Medizin von oben („top-down“) in die Versorgung implementieren lässt, wenn nicht eine kritische Masse an ärztlichen Entscheidungsträgern wenigstens über Grundkenntnisse in EbM verfügt. Allerdings sollte man vorsichtig sein, diese Defizite nur der Ärzteschaft anzulasten. Vielmehr sind sie auch die Konsequenz einer medizinischen Prioritätensetzung der Vergangenheit, in der ausschließlich die Grundlagenforschung gefördert worden war. Die internationale Entwicklung der klinischen Epidemiologie, fundamentale Voraussetzung von EbM, ging an Deutschland vorüber.

Perspektive

Damit parallel zu den Vorgaben des Gesetzgebers die Kultur einer evidenzbasierten Patientenversorgung entstehen kann, müsste nach dem Vorbild der angelsächsischen Länder eine evidenzbasierte Infrastruktur entwickelt werden.

- **Methodikwissenschaftler:** Die gerade begonnenen Veränderungen im Gesundheitssystem enthüllen einen erheblichen Mangel an Methodenwissenschaftlern und klinischen Epidemiologen, die den angestoßenen Prozess tragen, entwickeln und umsetzen. Weitere Magisterstudiengänge mit methodischem Schwerpunkt, wie z.B. klinische Epidemiologie und Versorgungsforschung (Health Care Research) werden dringend benötigt, um eine kritische Masse an Methodikern bereitzustellen. Dabei muss Sorge getragen werden, dass die klinische

Epidemiologie stärker in den medizinischen Fächern verwurzelt wird und nicht im Elfenbeinturm der theoretischen Fächer verschwindet.

- **Doppelqualifikation:** Möglichkeiten einer Doppelqualifikation in klinischer Medizin und klinischer Epidemiologie durch die Entwicklung von Rotationsprogrammen zwischen klinischer Epidemiologie und klinischer Medizin, wie z.B. nach dem Vorbild der „*Robert Woods Johnson Clinical Scholarship*“-Programmen in den USA (Kapitel 1), könnten diese Gefahr verhindern und die Ausbildung von zukünftigen Meinungsbildnern („Clinical Leadership“) sicherstellen.
- **Methodikforschung:** Um Anschluss an die internationale Diskussion zu gewinnen, sollte mindestens ein Zentrum für klinisch-epidemiologische (nicht biometrische) Methodikforschung aufgebaut werden, zusammen mit Möglichkeiten der öffentlichen Förderung einzelner Projekte.
- **Verbreitung des EbM-Curriculums der Ärztlichen Zentralstelle Qualitätssicherung und des Deutschen Netzwerks EbM:** Mit einer breit angelegten Förderung der Durchführung von EbM-Kursen nach dem EbM-Curriculum können die Inhalte der EbM und ihr Nutzen für die Patientenbetreuung in der Ärzteschaft angemessen verbreitet und Fehlinformationen und Missverständnisse ausgeräumt werden. Mittelfristig, wenn genügend Tutoren mit methodischer Qualifikation verfügbar sind, sollten die Inhalte und die Anwendung von EbM zu einem festen Bestandteil der Aus-, Weiter- und Fortbildung werden. In diesen Prozess sollten auch die nicht ärztlichen medizinischen Fachberufe einbezogen werden.
- **Bereitstellung von EbM-Ressourcen:** Beteiligung an den internationalen Bemühungen, die für EbM notwendigen Ressourcen an aufbereiteter Evidenz bereitzustellen, wie die Erstellung von Cochrane Reviews, Unterstützung der Produktion evidenzbasierter Informationsquellen und deren Verbreitung. Vorbildhaft dafür sind Internetentwicklungen wie die „National Electronic Library of Health“ (NELH) des NHS oder die kanadischen „Centres for Health Evidence“ (www.cche.net). Ein erster Schritt wäre die Bereitstellung einer nationalen Lizenz der Cochrane Library.

- **Institutionalisierung des Deutschen Cochrane Zentrums:** Nur durch eine langfristige solide Finanzierung von Institutionen wie dem Deutschen Cochrane Zentrum erhalten die bisher als „Graswurzel“-Aktivitäten organisierten Initiativen eine tragfähige Grundlage, um ihre Arbeit jenseits von Pilotprojekten weiterzuentwickeln und einen nennenswerten Beitrag zur Umsetzung der erklärten Prioritäten des Gesundheitssystems zu leisten.
- **Errichtung eines „Nationalen Zentrums“ für EbM:** Viele EbM-Aufgaben wie die Bereitstellung von hochwertiger Evidenz oder die Entwicklung und Durchführung von Schulungsprogrammen in größerem Rahmen sind so aufwändig, dass sie im lokalen und regionalen Rahmen als „Feierabend“-Projekte nicht oder nur schwer bewältigt werden können. Mit der Errichtung eines „nationalen EbM-Zentrums“, das überregionale Aufgaben übernimmt, lokale Initiativen unterstützt, Gremien berät und nationale und internationale Aktivitäten koordiniert, könnte die „Bottom-up“-Bewegung der evidenzbasierten Patientenversorgung gestärkt und die Ärzteschaft kompetent und effektiv mit den Methoden der evidenzbasierten Medizin vertraut gemacht werden.
- **Modellprojekte zur Implementierung von EbM:** Modellprojekte, die die inhaltlichen und strukturellen Voraussetzungen für eine patientenorientierte evidenzbasierte Medizin untersuchen und die Auswirkungen auf die Versorgung und die Gesundheit der Patienten erfassen, sollten unterstützt werden. Dies umfasst auch die Implementierung von Leitlinien und die Untersuchung ihrer Auswirkungen auf die Versorgung.

Für die Entscheidungen im Gesundheitssystem wie in der Patientenversorgung ist die klinische Epidemiologie / evidenzbasierte Medizin zu einer Realität geworden, an der Entscheidungsträger im System wie in der Versorgung nicht mehr vorbeikommen. Es wird höchste Zeit, die dafür notwendigen Grundlagen zu schaffen.

Literaturverzeichnis

1. Paul JR. Clinical Epidemiology. *J Clin Invest* 1938;17:539-41.
2. Feinstein AR. Clinical Epidemiology. I. The population experiments of nature and of men in human illness. *Ann Intern Med* 1968;69:807-20.
3. A.R. Feinstein. Clinical Epidemiology. The Architecture of Clinical Research. Philadelphia, London, Toronto: 1985.
4. Sackett DL. Clinical epidemiology. *Am.J Epidemiol.* 1969;89:125-8.
5. Fletcher RH. Clinical medicine meets modern epidemiology--and both profit. *Ann.Epidemiol.* 1992;2:325-33.
6. Cochrane, A. L. Effectiveness and Efficiency - Random Reflections on Health Services. The Nuffield Provincial Hospitals Trust. 1972.
7. A.L. Cochrane. 1931-1971: a critical review, with particular reference to the medical profession. *Medicines for the year 2000.*, pp 1-11. London: Office of Health Economics, 1979.
8. Oxford database of perinatal trials. Chalmers, I. Oxford disk issue 7(Version 1.2).1992.Oxford University Press.
9. Chalmers I, Enkin M, Keirse MJ. Effective care in pregnancy and childbirth. Oxford: Oxford University Press, 1989.
10. Enkin M, Keirse MJ, Chalmers I. A guide to effective care in pregnancy and childbirth. Oxford: Oxford University Press, 1989.
11. Shelley JH, Baur MP. Paul Martini: the first clinical pharmacologist? *Lancet* 1999;353:1870-3.
12. Martini P. Methodenlehre der Therapeutischen Untersuchung. Berlin: Springer Verlag, 1932.
13. Bleker J, Jachertz N, (Eds.). Medizin im "Dritten Reich". Cologne: Deutscher Aezzte-Verlag, 2002.
14. Bäumer A. NS-Biologie. Stuttgart: 1990.
15. Schwartz FW, Frhr.v.Troschke F, Wolter U. Public Health in Deutschland. In Public Health Forschungsverbände in der Deutschen Gesellschaft für Public Health e.V., ed. *Public Health Forschung in Deutschland.*, pp 23-32. Bern: Hans Huber, 1999.
16. Sackett DL. The competing objectives of randomized trials. *N.Engl.J Med* 1980;303:1059-60.
17. Chalmers TC, Celano P, Sacks HS, Smith H, Jr. Bias in treatment assignment in controlled clinical trials. *N.Engl.J.Med.* 1983;309:1358-61.
18. Sacks H, Chalmers TC, Smith H, Jr. Randomized versus historical controls for clinical trials. *Am.J.Med.* 1982;72:233-40.
19. Horwitz RI, Feinstein AR. Improved observational method for studying therapeutic efficacy. Suggestive evidence that lidocaine prophylaxis prevents death in acute myocardial infarction. *JAMA* 1981;246:2455-9.
20. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992;268:240-8.
21. Obias MD, Friedmann E, Brooks MM, Thomas SA, Haakenson C, Morris M *et al.* Adherence and arrhythmic mortality in the cardiac arrhythmia suppression trial (CAST). *Ann.Epidemiol.* 1996;6:93-101.
22. The EC/IC Bypass Study Group. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke; Results of an international randomized trial. *The new England Journal of Medicine* 1985;313:1191-200.

23. Besarab A, Bolton WK, Browne JK, Egrie JC, Nissenson AR, Okamoto DM *et al.* The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. *N.Engl.J.Med.* 1998;339:584-90.
24. Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B *et al.* Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* 1998;280:605-13.
25. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-33.
26. Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N.Engl.J Med* 2000;342:145-53.
27. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat.Med.* 1984; 3:409-22.
28. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. *Lancet* 1995;345:669-85.
29. Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. First International Study of Infarct Survival Collaborative Group. *Lancet* 1986;2:57-66.
30. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. *Lancet* 1988;2:349-60.
31. Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy--randomized trials in individual patients. *N.Engl.J Med* 1986;314:889-92.
32. Larson EB, Ellsworth AJ, Oas J. Randomized clinical trials in single patients during a 2-year period. *JAMA* 1993;270:2708-12.
33. Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. The implications of adopting a cluster design are still largely being ignored. *BMJ* 1998;317:1171-2.
34. Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979-1989. *Int J Epidemiol.* 1990;19:795-800.
35. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG, Donner A. Methods in health service research. Evaluation of health interventions at area and organisation level. *BMJ* 1999;319:376-9.
36. Gent M, Sackett DL. The qualification and disqualification of patients and events in long-term cardiovascular clinical trials. *Thromb.Haemost.* 1979;41:123-34.
37. Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N.Engl.J Med* 1979;301:1410-2.
38. Eisenberg JM. Clinical economics. A guide to the economic analysis of clinical practices. *JAMA* 1989;262:2879-86.
39. The EC/IC Bypass Study Group. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke; Results of an international randomized trial. *The new England Journal of Medicine* 1985;313:1191-200.
40. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ* 2002;324:1448-51.
41. Collaborative systematic review of the randomised trials of organised inpatient (stroke unit) care after stroke. Stroke Unit Trialists' Collaboration. *BMJ* 1997;314 :1151-9.
42. Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges LV *et al.* Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997;350:834-43.
43. Ernst E. Homöopathie: Argumente und Gegenargumente. *Deutsches Ärzteblatt* 1997;94:B-1910.

44. Ernst E, Pittler MH. Re-analysis of previous meta-analysis of clinical trials of homeopathy. *J Clin.Epidemiol.* 2000;53:1188.
45. Ernst E, White AR. Acupuncture for back pain: a meta-analysis of randomized controlled trials. *Arch.Intern.Med.* 1998;158:2235-41.
46. Irnich D, Behrens N, Molzen H, Konig A, Gleditsch J, Krauss M *et al.* Randomised trial of acupuncture compared with conventional massage and "sham" laser acupuncture for treatment of chronic neck pain. *BMJ* 2001;322:1574-8.
47. van Tulder MW, Cherkin DC, Berman B, Lao L, Koes BW. The effectiveness of acupuncture in the management of acute and chronic low back pain. A systematic review within the framework of the Cochrane Collaboration Back Review Group. *Spine* 1999;24:1113-23.
48. Ernst E. Herbal medicines: where is the evidence? *BMJ* 2000;321:395-6.
49. Effect of *Hypericum perforatum* (St John's wort) in major depressive disorder: a randomized controlled trial. *JAMA* 2002;287:1807-14.
50. Shelton RC, Keller MB, Gelenberg A, Dunner DL, Hirschfeld R, Thase ME *et al.* Effectiveness of St John's wort in major depression: a randomized controlled trial. *JAMA* 2001;285:1978-86.
51. Ernst E, Assendelft WJ. Chiropractic for low back pain. We don't know whether it does more good than harm. *BMJ* 1998;317:160.
52. Ottenbacher KJ, Jannell S. The results of clinical trials in stroke rehabilitation research. *Arch Neurol.* 1993;50:37-44.
53. Fontanarosa PB, Lundberg GD. Alternative medicine meets science. *JAMA* 1998;280:1618-9.
54. Tonelli MR, Callahan TC. Why alternative medicine cannot be evidence-based. *Acad Med* 2001;76:1213-20.
55. Guyatt GH, Rennie D. Users' guides to the medical literature. *JAMA* 1993;270:2096-7.
56. Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Eval.Health Prof.* 2002;25:12-37.
57. Spitzer WO. Meta-meta-analysis: unanswered questions about aggregating data. *J Clin.Epidemiol.* 1991;44:103-7.
58. Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol.* 1994;140:771-8.
59. Spitzer WO. The challenge of meta-analysis. *J Clin.Epidemiol.* 1995;48:1-4.
60. Bailar J. The Promise and Problems of Meta-Analysis. *The new England Journal of Medicine* 1997;337:559-61.
61. Meta-analysis under scrutiny. *Lancet* 1997;350:675.
62. Bucher HC, Guyatt GH, Cook RJ, Hatala R, Cook DJ, Lang JD *et al.* Effect of calcium supplementation on pregnancy-induced hypertension and preeclampsia: a meta-analysis of randomized controlled trials. *JAMA* 1996;275:1113-7.
63. Horner SM. Efficacy of intravenous magnesium in acute myocardial infarction in reducing arrhythmias and mortality. Meta-analysis of magnesium in acute myocardial infarction. *Circulation* 1992;86:774-9.
64. Teo KK, Yusuf S, Collins R, Held PH, Peto R. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. *BMJ* 1991;303:1499-503.
65. Yusuf S, Teo K, Woods K. Intravenous magnesium in acute myocardial infarction. An effective, safe, simple, and inexpensive intervention. *Circulation* 1993;87:2043-6.
66. Levine RJ, Hauth JC, Curet LB, Sibai BM, Catalano PM, Morris CD *et al.* Trial of calcium to prevent preeclampsia. *N.Engl.J.Med.* 1997;337:69-76.
67. Egger M, Smith GD. Misleading meta-analysis. *BMJ.* 1995;310:752-4.

68. Egger M, Zellweger ZT, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German . *Lancet* 1997;350:326-9.
69. Juni P, Hoenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol.* 2002;31:115-23.
70. Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test . *BMJ* 1997;315:629-34.
71. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin.Epidemiol.* 2001;54:1046-55.
72. Simes RJ. Publication bias: the case for an international registry of clinical trials. *J.Clin.Oncol.* 1986;4:1529-41.
73. European Medical Research Councils (EMRC). Controlled clinical trials. 13, 1-6. 2001. European Science Foundation (ESF).
74. Tramer MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997;315:635-40.
75. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control.Clin.Trials.* 1995;16:62-73.
76. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis . *JAMA* 1999;282:1054-60.
77. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D *et al.* The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
78. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I *et al.* Improving the quality of reporting of randomized controlled trials. The CONSORT statement . *JAMA* 1996;276:637-9.
79. Clarke M. The QUORUM statement. *Lancet* 2000;355:756-7.
80. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D *et al.* Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008-12.
81. Memorandum zur Evaluierung diagnostischer Maßnahmen. In Köbberling J, Trampisch HJ, Windeler J, eds. Schattauer, 1989.
82. Sheps SB, Schechter MT. The assessment of diagnostic tests. A survey of current medical research. *JAMA* 1984;252:2418-22.
83. Arroll B, Schechter MT, Sheps SB. The assessment of diagnostic tests: a comparison of medical literature in 1982 and 1985. *J Gen.Intern Med* 1988;3:443-7.
84. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995;274:645-51.
85. Simel DL, Feussner JR, DeLong ER, Matchar DB. Intermediate, indeterminate, and uninterpretable diagnostic test results. *Med Decis.Making* 1987;7:107-14.
86. Lijmer JG, Mol BW, Heisterkamp S, Bossuyt GJ, Prins MH, van der Meulen JH *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
87. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat.Med.* 2002;21:1525-37.
88. Feinstein AR. Misguided efforts and future challenges for research on "diagnostic tests" . *J Epidemiol.Community Health* 2002;56:330-2.
89. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC *et al.* Guidelines for meta-analyses evaluating diagnostic tests . *Ann.Intern.Med.* 1994;120:667-76.

90. Midgette AS, Stukel TA, Littenberg B. A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. *Med.Decis.Making*. 1993;13:253-7.
91. Shapiro DE. Issues in combining independent estimates of the sensitivity and specificity of a diagnostic test. *Acad.Radiol*. 1995;2 Suppl 1:S37-S47.
92. Walter SD, Irwig L, Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. *J.Clin.Epidemiol*. 1999;52:943-51.
93. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J.Clin.Epidemiol*. 1995;48:119-30.
94. Perrier A, Desmarais S, Miron MJ, de MP, Lepage R, Slosman D *et al*. Non-invasive diagnosis of venous thromboembolism in outpatients. *Lancet* 1999;353:190-5.
95. Prandoni P, Lensing AW, Cogo A, Cuppini S, Villalta S, Carta M *et al*. The long-term clinical course of acute deep venous thrombosis . *Ann Intern Med*. 1996;125:1-7.
96. Anis AH, Stiell IG, Stewart DG, Laupacis A. Cost-effectiveness analysis of the Ottawa Ankle Rules. *Ann Emerg.Med* 1995;26:422-8.
97. Auleley GR, Ravaud P, Giraudeau B, Kerboull L, Nizard R, Massin P *et al*. Implementation of the Ottawa ankle rules in France. A multicenter randomized controlled trial . *JAMA* 1997;277:1935-9.
98. Stiell I, Wells G, Laupacis A, Brison R, Verbeek R, Vandemheen K *et al*. Multicentre trial to introduce the Ottawa ankle rules for use of radiography in acute ankle injuries. Multicentre Ankle Rule Study Group. *BMJ* 1995;311:594-7.
99. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology. A basic science for clinical medicine*. Boston, Toronto, London: Little, Brown and Company, 1991.
100. Raspe HH. Evidence based medicine: Modischer Unsinn, alter Wein in neuen Schlauchen oder aktuelle Notwendigkeit? *Z.Arztl.Fortbild.Jena*. 1996;90:553-62.
101. Wegscheider K. Evidenzbasierte Medizin - Irrweg oder Aufbruch zu neuen Ufern? *Herzschr Elektrophys* 2000;11:69-76.
102. How to read clinical journals: I. why to read them and how to start reading them critically. *Can Med Assoc J* 1981;124:555-8.
103. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology: The Essentials*. Williams Wilkins, 1996.
104. Sackett DL, Haynes RB, Tugwell P, Guyatt GH. *Clinical Epidemiology : A Basic Science for Clinical Medicine*. Boston, Toronto, London: Little Brown and Company. 1991, 1985.
105. Maynard A. Evidence-based medicine: an incomplete method for informing treatment choices . *Lancet* 1997;349:126-8.
106. Miles A, Charlton B, Bentley P, Polychronis A, Grey J, Price N. New perspectives in the evidence-based healthcare debate. *J Eval.Clin.Pract*. 2000;6:77-84.
107. Naylor CD. Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet* 1995;345:840-2.
108. Buetow S., Kenealy T. Evidence-based medicine: the need for a new definition. *J Eval.Clin.Pract*. 2000;6:85-92.
109. Rogler G, Schrader H. Evidence-based Medicine - oder: Die trügerische Sicherheit der Evidenz. *DMW* 20010;125:1122-8.
110. Straus SE, McAlister FA. Evidence-based medicine: a commentary on common criticisms. *CMAJ*. 2000;163:837-41.
111. Raspe H. Ethical implications of evidence-based medicine. *Dtsch.Med Wochenschr*. 2002;127:1769-73.
112. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't . *BMJ*. 1996;312:71-2.

113. Mühlhauser I, Höldke B. Mammographie-Screening - Darstellung der wissenschaftlichen Evidenzgrundlage zur Kommunikation mit der Frau. *arznei-telegramm* 1999;10 :101.
114. Wehner M, Köbberling J. Strategien zur Umstrukturierung einer Klinik in Richtung auf EBM. In Klotz T, Grüne F, Weigand C, Lauterbach K, Schrappe M, eds. *Gesundheitspolitik, Krankenhausorganisation, Qualitätsmanagement: 3. Kölner Krankenhaus-Kongress*, pp 191-205. Göttingen: Cuviller Verlag., 2000.
115. Zielinski W, Goldstein M, König U. Evidence-based medicine in internal guideline development in a general hospital--the Park-Clinic EbM-Project. *Z.Arztl.Fortbild.Qualitatssich.* 2001;95:413-7.
116. Ärztliche Zentralstelle Qualitätssicherung. Leitlinien-Clearing-Bericht "Hypertonie". Germering/München: W.Zuckschwerdt Verlag, 2000.
117. Developing a Methodology for Drawing up Guidelines on Best Medical Practice (Recommendations Rec(2001)13. 2002. Strasbourg, Council of Europe Publishing.
118. Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1992;102:305S-11S.
119. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH *et al.* Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N.Engl.J.Med* 1991;324:781-8.
120. Valenstein ES. Great and Desperate Cures: The Rise and Decline of Psychosurgery and other Radical Treatments for Mental Illness. p 338. New York: Basic Books, 1986.
121. Chalmers TC, Smith H, Jr., Blackburn B, Silverman B, Schroeder B, Reitman D *et al.* A method for assessing the quality of a randomized control trial. *Control Clin.Trials* 1981;2:31-49.
122. Byar DP. Problems with using observational databases to compare treatments. *Stat.Med.* 1991;10:663-6.
123. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs . *N.Engl.J.Med.* 2000;342:1887-92.
124. Hlatky MA, Califf RM, Harrell FE, Jr., Lee KL, Mark DB, Pryor DB. Comparison of predictions based on observational data with the results of randomized controlled clinical trials of coronary artery bypass surgery. *J.Am.Coll.Cardiol.* 1988;11:237-45.
125. Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am.J.Med.* 1990;89:630-8.
126. Horwitz RI, Feinstein AR. Improved observational method for studying therapeutic efficacy. Suggestive evidence that lidocaine prophylaxis prevents death in acute myocardial infarction. *JAMA* 1981;246:2455-9.
127. Chalmers TC, Matta RJ, Smith H, Jr., Kunzler AM. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N.Engl.J.Med.* 1977;297:1091-6.
128. Sacks H, Chalmers TC, Smith H, Jr. Randomized versus historical controls for clinical trials. *Am.J.Med.* 1982;72:233-40.
129. Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Control Clin.Trials* 1990;11:339-52.
130. Reimold SC, Chalmers TC, Berlin JA, Antman EM. Assessment of the efficacy and safety of antiarrhythmic therapy for chronic atrial fibrillation: observations on the role of trial design and implications of drug-related mortality. *Am.Heart J.* 1992;124:924-32.
131. Diehl LF, Perry DJ. A comparison of randomized concurrent control groups with matched historical control groups: are historical controls valid? *J.Clin.Oncol.* 1986;4:1114-20.
132. Worldwide collaborative observational study and meta-analysis on allogenic leukocyte immunotherapy for recurrent spontaneous abortion. Recurrent Miscarriage Immunotherapy Trialists Group. *Am.J.Reprod.Immunol.* 1994;32:55-72.

133. Watson A, Vandekerckhove P, Lilford R, Vail A, Brosens I, Hughes E. A meta-analysis of the therapeutic role of oil soluble contrast media at hysterosalpingography: a surprising result? *Fertil.Steril.* 1994;61:470-7.
134. Carroll D, Tramer M, McQuay H, Nye B, Moore A. Randomization is important in studies with pain outcomes: systematic review of transcutaneous electrical nerve stimulation in acute postoperative pain. *Br.J.Anaesth.* 1996;77:798-803.
135. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: Medical. *Stat.Med.* 1989;8:441-54.
136. Miller JN, Colditz GA, Mosteller F. How study design affects outcomes in comparisons of therapy. II: Surgical. *Stat.Med* 1989;8:455-66.
137. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
138. Ottenbacher K. Impact of random assignment on study outcome: an empirical examination. *Control Clin.Trials* 1992;13:50-61.
139. Imperiale TF, McCullough AJ. Do corticosteroids reduce mortality from alcoholic hepatitis? A meta-analysis of the randomized trials. *Ann.Intern.Med.* 1990;113:299-307.
140. Nurmohamed MT, Rosendaal FR, Buller HR, Dekker E, Hommes DW, Vandembroucke JP *et al.* Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet* 1992;340:152-6.
141. Khan KS, Daya S, Jadad A. The importance of quality of primary studies in producing unbiased systematic reviews. *Arch.Intern.Med.* 1996;156:661-6.
142. Ortiz Z, Shea B, Suarez-Almazor ME, Moher D, Wells GA, Tugwell P. The efficacy of folic acid and folinic acid in reducing methotrexate gastrointestinal toxicity in rheumatoid arthritis. A metaanalysis of randomized controlled trials. *J.Rheumatol.* 1998;25:36-43.
143. Pyorala S, Huttunen NP, Uhari M. A review and meta-analysis of hormonal treatment of cryptorchidism. *J.Clin.Endocrinol.Metab* 1995;80:2795-9.
144. Pyorala K, Pedersen TR, Kjekshus J, Faergeman O, Olsson AG, Thorgeirsson G. Cholesterol lowering with simvastatin improves prognosis of diabetic patients with coronary heart disease. A subgroup analysis of the Scandinavian Simvastatin Survival Study (4S). *Diabetes Care* 1997;20:614-20.
145. Khan KS, Daya S, Collins JA, Walter SD. Empirical evidence of bias in infertility research: overestimation of treatment effect in crossover trials using pregnancy as the outcome measure. *Fertil.Steril.* 1996;65:939-45.
146. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG *et al.* Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;286:821-30.
147. Kunz, R., Vist, G., and Oxman, A. D. Randomisation to protect against selection bias in healthcare trials. The Cochrane Library (3). 2002.
148. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol.Assess.* 1998;2:i-124.
149. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999;319:312-5.
150. Djulbegovic, B. Two systematic reviews-two different answers. www.bmj.com/cgi/eletters/319/7205/312;18.08.1999
151. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.
152. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N.Engl.J.Med.* 2000;342:1878-86.

153. Kunz R, Khan KS, Neumayer HH. Observational studies and randomized trials. *N.Engl.J Med* 2000;343:1194-5.
154. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N.Engl.J Med* 2000;342:1907-9
155. Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B *et al.* Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* 1998;280:605-13.
156. Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N.Engl.J Med* 2000;342:145-53.
157. Omenn GS, Goodman GE, Thornquist MD, Balmes J, *et al.* Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *N.Engl.J Med* 1996;334:1150-5.
158. Kunz R, Khan KS, Neumayer HH. Observational studies and randomized trials. *N.Engl.J Med* 2000;343:1194-5.
159. Khan KS, Kunz R, Kleijnen J, Antes G. Systematic Reviews to Support Evidence-based Medicine: How to review and apply findings of health care literature. London: The Royal Soc Med Press, 2002.
160. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.
161. Stables RH. Observational research in the evidence based environment: eclipsed by the randomised controlled trial? *Heart* 2002;87:101-2.
162. Elphick HE, Tan A, Ashby D, Smyth RL. Systematic reviews and lifelong diseases. *BMJ* 2002;325:381-4.
163. Kunz, R., Cowan, D., Halpern, S., Oxman, A. D., and Sackett, D. L. Development of a database for empirical studies on methodological issues. 2nd International Cochrane Meeting Hamilton, Canada. 1994.
164. Cook DJ, Reeve BK, Guyatt GH, Heyland DK, Griffith LE, Buckingham L *et al.* Stress ulcer prophylaxis in critically ill patients. Resolving discordant meta-analyses. *JAMA* 1996;275:308-14.
165. Sackett DL, Deeks JJ, Douglas GA. Down with odds ratios! *Evidence-Based Medicine* 1996;1:164.
166. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N.Engl.J Med.* 1988;318:1728-33.
167. Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, MacFarlane PW *et al.* Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. *N.Engl.J Med* 1995; 333:1301-7.
168. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ* 1995;311:1356-9.
169. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ.* 1995;310:452-4.
170. Tversky A, Kahnemann D. Judgement under Uncertainty:heuristics and Biases. *Science* 1974;185:1124-31.
171. Kahneman D, Tversky A. Choices, Values and Frames. *American Psychologist* 1984;39:341-50.
172. Gigerenzer G. We need statistical thinking, not statistical rituals. *Behavioral And Brain Sciences* 1998;21:199-200.
173. Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Medicine. Communicating statistical information. *Science* 2000;290:2261-2.
174. Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *BMJ.* 1994;309:761-4.
175. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann.Intern.Med.* 1992;117:916-21.
176. Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med* 1992;92:121-4.

177. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994;343:1209-11.
178. Cranney M, Walley T. Same information, different decisions: the influence of evidence on the management of hypertension in the elderly. *Br J Gen.Pract.* 1996;46:661-3.
179. Friedmann PD, Brett AS, Mayo-Smith MF. Differences in generalists' and cardiologists' perceptions of cardiovascular risk and the outcomes of preventive therapy in cardiovascular disease. *Ann.Intern.Med.* 1996;124:414-21.
180. Grover SA, Lowensteyn I, Esrey KL, Steinert Y, Joseph L, Abrahamowicz M. Do doctors accurately assess coronary risk in their patients? Preliminary results of the coronary health assessment study. *BMJ* 1995;310:975-8.
181. Hamilton-Craig CR, Tonkin AL, Jobling RG. How accurate are hypertension treatment decisions? Absolute risk assessment and prescribing for moderate hypertension--a study of Cambridge general practitioners. *Blood Press* 2000;9:323-7.
182. Randolph AG, Haynes RB, Wyatt JC, Cook DJ, Guyatt GH. Users' Guides to the Medical Literature: XVIII. How to use an article evaluating the clinical impact of a computer-based clinical decision support system. *JAMA* 1999;282:67-74.
183. Steel N. Thresholds for taking antihypertensive drugs in different professional and lay groups: questionnaire survey. *BMJ* 2000;320:1446-7.
184. Misselbrook D, Armstrong D. Patients' responses to risk information about the benefits of treating hypertension. *Br J Gen.Pract.* 2001;51:276-9.
185. Rajagopal S, Goodman PJ, Tannock IF. Adjuvant chemotherapy for breast cancer: discordance between physicians' perception of benefit and the results of clinical trials. *J Clin.Oncol.* 1994;12:1296-304.
186. Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen. Bedarfsgerechtigkeit und Wirtschaftlichkeit - Band III Über-, Unter- und Fehlversorgung. 2001.
187. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.
188. Cook DJ, Fuller HD, Guyatt GH, Marshall JC, Leasa D, Hall R *et al.* Risk factors for gastrointestinal bleeding in critically ill patients. Canadian Critical Care Trials Group. *N.Engl.J.Med.* 1994;330:377-81.
189. Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *Int J Epidemiol.* 2002;31:72-6.
190. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA* 1995;274:1800-4.
191. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
192. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M *et al.* Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-13.
193. Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998;317:1309-12.
194. Grodstein F, Stampfer MJ, Manson JE, Colditz GA, Willett WC, Rosner B *et al.* Postmenopausal estrogen and progestin use and the risk of cardiovascular disease. *N.Engl.J.Med.* 1996;335:453-61.
195. The Evidence-Based Medicine Working Group. Users' Guides to the medical literature - A Manual for Evidence-Based Clinical Practice. JAMA & Archives Journals, 2002.
196. Bundesärztekammer KB. Das Leitlinien-Clearingverfahren - Ziele und Arbeitsplan. *Deutsches Ärzteblatt* 1999;96:A-2105-A-2106.
197. Shaneyfelt T, Mayo-Smith M, Rothwangl J. Are Guidelines Following Guidelines? The Methodological Quality of Clinical Practice Guidelines in the Peer-Reviewed Medical Literature. *JAMA* 1999;281:1900-5.

198. Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999;11:21-8.
199. Helou A, Perleth M, Schwartz FW. Determining priorities in the development of medical guidelines. 1: Criteria, procedures and actors. *Z Arztl Fortbild Qualitatssich* 2000;94:53-60.
200. Deutsche Liga zur Bekämpfung des hohen Blutdrucks e.V. Arterielle Hypertonie.1998.www.uni-duesseldorf.de/WWW/AWMF/II/ihyp0.1.htm
201. Arzneimittelkommission der Deutschen Ärzteschaft. Empfehlungen zur Therapie der arteriellen Hypertonie. *Arzneiverordnungen in der Praxis*. 1998;Sonderheft 9.
202. Feldman R, Campbell N, Larochelle P, for the Task Force of the 1999 Canadian Recommendations for the Management of Hypertension. Recommendations for the Management of Hypertension. *CMAJ*. 1999;161:S1-ff.
203. Canadian Coalition for High Blood Pressure Prevention and Control. Lifestyle Modification to prevent and control hypertension. *CMAJ*. 1999;1999:S1-S50.
204. National Institute of Health. The 6th Report of the Joint National Committee on the Prevention, Detection, Evaluation and Treatment of High Blood Pressure. NIH-Publ.98-4080.1997. www.nhlbi.nih.gov/guidelines/hypertension/jncintro.htm
205. Institute for Clinical System Integration (ICSI). Hypertension -Diagnosis and Treatment. Bloomington (MN): Ardel Group, 1999.
206. British Hypertension Society. Guidelines for the Management of Hypertension. *J Hum Hypertension* 1999;13:569-92.
207. Hypertension Society of South Africa. Guidelines for the Management of hypertension at primary health level. *S.Afr.Med J* 1995;95:1321-5.
208. Department of Veterans Affairs. The Pharmacological Management of Hypertension. U.S.Veterans Health Administration Publ.-No. 96-0003 12/1996. 2002. www.va.gov/health/hypertension/HTN.doc
209. Chalmers JealftW-IHGC. 1999 Guidelines for the Management of Hypertension. *J Hypertens*. 1999;17:151-85.
210. New Zealand Guidelines Group. Guidelines for the Management of mildly raised blood pressure in New Zealand. National Health Committee.2002. www.nzgg.org.nz/library/gl_complete/blood_pressure/index.htm
211. Erdmann E. Missing medical care? *Dtsch.Med Wochenschr*. 2002;127:1741.
212. Ramsey PG, Carline JD, Inui TS, Larson EB, LoGerfo JP, Norcini JJ *et al*. Changes over time in the knowledge base of practicing internists. *JAMA* 1991;266:1103-7.
213. Leigh TM, Young PR, Haley JV. Performances of family practice diplomates on successive mandatory recertification examinations. *Acad.Med* 1993;68:912-9.
214. Caulford PG, Lamb SB, Kaigas TB, Hanna E, Norman GR, Davis DA. Physician incompetence: specific problems and predictors. *Acad.Med* 1994;69:S16-S18.
215. Osheroff JA, Forsythe DE, Buchanan BG, Bankowitz RA, Blumenfeld BH, Miller RA. Physicians' information needs: analysis of questions posed during clinical teaching. *Ann Intern Med* 1991;114:576-81.
216. Gorman PN, Helfand M. Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered. *Med Decis.Making* 1995;15:113-9.
217. Williamson JW, German PS, Weiss R, Skinner EA, Bowes F, III. Health science information management and continuing education of physicians. A survey of U.S. primary care practitioners and their opinion leaders. *Ann.Intern Med* 1989;110:151-60.
218. Thomson O'Brien MA, Freemantle N, Oxman AD, Wolf F, Davis DA, Herrin J. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane.Database.Syst.Rev*. 2001;CD003030.

219. Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *JAMA* 1999;282:867-74.
220. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. *JAMA* 1995;274:700-5.
221. Gerst Th. Ärztlicher Kompetenzerhalt ist Sache der Selbstverwaltung. *Deutsches Ärzteblatt* 2002;A 1940.
222. Evidence-Based Care Resource Group. Evidence-based care: 5. Lifelong learning: how can we learn to be more effective? *CMAJ*. 1994;150:1971-3.
223. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine - How to Practice and Teach EBM. Churchill Livingstone, 2000.
224. The Evidence-Based Medicine Working Group. User's Guides to the Medical Literature - A Manual for Evidence-Based Clinical Practice. *JAMA & Archives Journals*, 2001.
225. Shaughnessy AF, Slawson DC. Are we providing doctors with the training and tools for lifelong learning? *BMJ* 1999;319:1280.
226. Burls, A. J. E. An evaluation of the impact of half day workshops teaching critical appraisal skills. 1997. Oxford, Institute of Health Sciences.
227. Fritsche L., Neumayer H.-H., and Kunz, R. Vom Konzept in die Praxis. Umsetzbarkeit eines strukturierten Programms für einen Grundkurs in Evidence-based Medicine. *ZaeFQ* 93, 73-74. 1999.
228. Smith CA, Ganschow PS, Reilly BM, Evans AT, *et al.* Teaching residents evidence-based medicine skills: a controlled trial of effectiveness and assessment of durability. *J Gen. Intern Med* 2000;15:710-5.
229. Sackett DL, Parkes J. Teaching critical appraisal: no quick fixes. *CMAJ*. 1998;158:203-4.
230. Fritsche, L., Greenhalgh, T., Falck-Ytter, Y., Neumayer H.-H., and Kunz, R. Do short courses in evidence-based medicine improve knowledge and skills? Validation of the Berlin Questionnaire and Before and After Study of EBM training. *BMJ* (accepted) . 2002.
231. Norman GR, Shannon SI. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *CMAJ*. 1998;158:177-81.
232. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-5.
233. The Cochrane Library, Issue 3. 1999. Oxford, Update Software.
234. Best Evidence 3. 1999. American College of Physicians.
235. Clinical Evidence. Bern, Göttingen, Toronto, Seattle: Verlag Hans Huber, 1999.
236. Guyatt GH, Meade MO, Jaeschke RZ, Cook DJ, Haynes RB. Practitioners of evidence based care. Not all clinicians need to appraise evidence from scratch but all need some skills. *BMJ* 2000;320:954-5.
237. McColl A, Smith H, White P, Field J. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. *BMJ* 1998;316:361-5.
238. Green ML. Evidence-based medicine training in graduate medical education: past, present and future. *J Eval. Clin. Pract.* 2000;6:121-38.
239. Taylor DK, Buterakos J. Evidence-based medicine: not as simple as it seems. *Acad. Med.* 1998;73:1221-2.
240. Abrahamson S, Baron J, Elstein AS, Hammond WP, Holzman GB, Marlow B *et al.* Continuing medical education for life: eight principles. *Acad. Med.* 1999;74:1288-94.
241. Dobbie AE, Schneider FD, Anderson AD, Littlefield J. What evidence supports teaching evidence-based medicine? *Acad Med* 2000;75:1184-5.
242. Audet N, Gagnon R, Ladouceur R, Marcil M. How effective is the teaching of critical analysis of scientific publications? Review of studies and their methodological quality. *CMAJ*. 1993;148:945-52.
243. Norman GR. Examining the assumptions of evidence-based medicine. *J. Eval. Clin. Pract.* 1999;5:139-47.

244. Green ML. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. *Acad Med* 1999;74:686-94.
245. Parkes, J., Hyde, C., Deeks, J., and Milne, R. Teaching critical appraisal skills in health care settings (Cochrane Review). In: The Cochrane Library (3). 2001. Oxford, Update Software.
246. Linzer M, Brown JT, Frazier LM, DeLong ER, Siegel WC. Impact of a medical journal club on house-staff reading habits, knowledge, and critical appraisal skills. A randomized control trial. *JAMA* 1988;260:2537-41.
247. Kirkpatrick DI. Evaluation of training. In Craig R.Bittel I., ed. *Training and Development handbook*, New York: McGraw-Hill, 1967.
248. Hutchinson L. Evaluating and researching the effectiveness of educational interventions. *BMJ* 1999;318:1267-9.
249. Streiner.D.L., Norman GR. Health Measurement Scales. Oxford: University Press, 1995.
250. Stern DT, Linzer M, O'Sullivan PS, Weld L. Evaluating medical residents' literature-appraisal skills. *Acad Med* 1995;70:152-4.
251. Norman GR, Van Der Vleuten CPM, Newble D. International Handbook of Research in Medical Education. Kluwer Academic Publishers, 2002.
252. Wilkes M, Bligh J. Evaluating educational interventions. *BMJ* 1999;318:1269-72.
253. Rogler G, Schrader H. Evidence-based Medicine - oder: Die trügerische Sicherheit der Evidenz. *DMW* 2001;125:1122-8.
254. Sackett DL, Richardson WS, Rosenberg W, Haynes RB, (Deutsche Ausgabe: R.Kunz und L.Fritsche). Evidenzbasierte Medizin. EBM-Umsetzung und -Vermittlung. München: Zuckschwerdt Verlag, 1999.
255. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. *JAMA* 1993;270:2093-5.
256. Curriculum Evidenzbasierte Medizin. Ärztliche Zentralstelle Qualitätssicherung und Deutsches Netzwerk EbM. DNEbM, 1-50. 2002. 19-9-0002. www.ebm-netzwerk.de
257. Kunz, R., Fritsche L., and Neumayer H.-H. Quality assurance criteria for continuing education in evidence-based medicine. *ZaeFQ* 95, 371-375. 2001.
258. Greenhalgh T. Consensus Research. In Elwyn G, Greenhalgh T, Macfarlane F, eds. *Groups - a Guide to Small Group Work in Health Care Management, Education and Research.*, Oxford: Radcliffe Medical Press, 2000.
259. Lienert GA, Raatz U. Testaufbau und Testanalyse. Weinheim: Psychologie Verlagsunion, 1995.
260. Nunnally JC. Psychometric Theory. New York: McGraw-Hill, 1978.
261. Cronbach LJ. Coefficient alpha and the Internal Structure of Tests. *Psychometrika* 1951;16:297-334.
262. Khan KS, Awonuga AO, Dwarakanath LS, Taylor R. Assessments in evidence-based medicine workshops: loose connection between perception of knowledge and its objective assessment. *Med Teach* 2001;23:92-4.
263. Knowles MS. Modern Practice of Adult Education - From Pedagogy to Andragogy. Chicago: Follett Publishing, 1984.
264. Green ML. Evidence-based medicine training in internal medicine residency programs a national survey. *J Gen. Intern Med* 2000;15:129-33.
265. Greenhalgh T, Douglas HR. Experiences of general practitioners and practice nurses of training courses in evidence-based health care: a qualitative study. *Br.J.Gen.Pract.* 1999;49:536-40.
266. Freeman AC, Sweeney K. Why general practitioners do not implement evidence: qualitative study. *BMJ* 2001;323:1100-2.
267. Cranney M, Walley T. Same information, different decisions: the influence of evidence on the management of hypertension in the elderly. *Br J Gen.Pract.* 1996;46:661-3.

268. Hersh WR, Hickam DH. How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. *JAMA* 1998;280:1347-52.
269. Hunt DL, Jaeschke R, McKibbin KA. Users' guides to the medical literature: XXI. Using electronic health information resources in evidence-based practice. Evidence-Based Medicine Working Group. *JAMA* 2000;283:1875-9.
270. Jadad AR, Gagliardi A. Rating health information on the Internet: navigating to knowledge or to Babel? *JAMA* 1998;279:611-4.
271. Oakley A, Rajan L, Robertson P. A comparison of different sources of information about pregnancy and childbirth. *J.Biosoc.Sci.* 1990;22:477-87.
272. Fairhurst K, Huby G. From trial data to practical knowledge: qualitative study of how general practitioners have accessed and used evidence about statin drugs in their management of hypercholesterolaemia. *BMJ* 1998;317:1130-4.
273. Tomlin Z, Humphrey C, Rogers S. General practitioners' perceptions of effective health care. *BMJ* 1999;318:1532-5.
274. Timpka T, Ekstrom M, Bjurulf P. Information needs and information seeking behaviour in primary health care. *Scand.J Prim.Health Care* 1989;7:105-9.
275. Wyatt JC. Clinical questions and information needs. *J R.Soc.Med* 2000;93:168-71.
276. Green ML, Ciampi MA, Ellis PJ. Residents' medical information needs in clinic: are they being met? *Am J Med* 2000;109:218-23.
277. Davidoff F, Florance V. The informationist: a new health profession? *Ann.Intern.Med.* 2000;132:996-8.
278. Horton R. The information wars. *Lancet* 1999;353:164-5.
279. Chassin MR. Is health care ready for Six Sigma quality? *Milbank Q.* 1998;76:565-91, 510.
280. Grol R. Improving the quality of medical care: building bridges among professional pride, payer profit, and patient satisfaction. *JAMA* 2001;286:2578-85.
281. Lomas J, Anderson GM, Domnick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N.Engl.J Med* 1989;321:1306-11.
282. van der Weijden T, Grol R. Preventing recurrent coronary heart disease. We need to attend more to implementing evidence based practice. *BMJ* 1998;316:1400-1.
283. Donald, A. The front-line evidence-based medicine project. 1998. NHS Executive North Thames regional office research and development.
284. Dunning, M, Abi-Aad, G, Gilbert, D, Hutton, H, and Brown, C. Experience, Evidence and Everyday Practice: Creating systems for delivering effective health care. 1999. King's Fund.
285. Feder G, Griffiths C, Eldridge S, Spence M. Effect of postal prompts to patients and general practitioners on the quality of primary care after a coronary event (POST): randomised controlled trial. *BMJ* 1999;318:1522-6.
286. Cupples ME, McKnight A. Randomised controlled trial of health promotion in general practice for patients at high cardiovascular risk. *BMJ* 1994;309:993-6.
287. Christakis DA, Zimmerman FJ, Wright JA, Garrison MM, Rivara FP, Davis RL. A randomized controlled trial of point-of-care evidence to improve the antibiotic prescribing practices for otitis media in children. *Pediatrics* 2001;107:E15.
288. Balas EA, Boren SA, Hicks LL, Chonko AM, Stephenson K. Effect of linking practice data to published evidence. A randomized controlled trial of clinical direct reports. *Med.Care* 1998;36:79-87.
289. Campbell NC, Thain J, Deans HG, Ritchie LD, Rawles JM, Squair JL. Secondary prevention clinics for coronary heart disease: randomised trial of effect on health. *BMJ* 1998;316:1434-7.

290. McAlister FA, Lawson FM, Teo KK, Armstrong PW. Randomised trials of secondary prevention programmes in coronary heart disease: systematic review. *BMJ* 2001;323:957-62.
291. Busse, R., Welte, T., Bode-Böger, St., and Böger, R. Kommunikationsprozess und Therapie-Entscheidungen an der Schnittstelle von stationärer und ambulanter Versorgung im West-/ Ostvergleich. Norddeutscher Forschungsverbund Public Health. Endbericht Projekt D II 1. 1999. Hannover.
292. von Luckner AG. Medical correspondence. *Z.Arztl.Fortbild.Qualitatssich.* 1997;91:563-7.
293. Newton J, Eccles M, Hutchinson A. Communication between general practitioners and consultants: what should their letters contain? *BMJ* 1992;304:821-4.
294. Isaac DR, Gijsbers AJ, Wyman KT, Martyres RF, Garrow BA. The GP-hospital interface: attitudes of general practitioners to tertiary teaching hospitals. *Med J Aust.* 1997;166:9-12.
295. Himmel W, Tabache M, Kochen MM. What happens to long-term medication when general practice patients are referred to hospital? *Eur.J Clin.Pharmacol.* 1996;50:253-7.
296. McKinlay JB. More appropriate evaluation methods for community-level health interventions. Introduction to the special issue. *Eval.Rev* 1996;20:237-43.
297. Donner A. Sample size requirement for stratified cluster randomization designs. *Stat.Med* 1992;11:743-50
298. Cates C, Senn S. Screening mammography re-evaluated. *Lancet* 2000;355:750.
299. Senn S. Screening for breast cancer with mammography. *Lancet* 2001;358:2165.
300. Zielinski W, Kunz R, Rakowsky N, Müller-Lissner S. Niedergelassene Ärzte zur Qualität von Arztbriefen. Ergebnisse einer Befragung aus dem Park-Klinik EbM-Projekt. Projektbaustein EbM-Handlungsempfehlungen. *Gesundheitsökonomie & Qualitätsmanagement* 2002;7:225-31.
301. Stalhammar J, Holmberg L, Svardsudd K, Tibblin G. Written communication from specialists to general practitioners in cancer care. What are the expectations and how are they met? *Scand.J Prim.Health Care* 1998;16:154-9.
302. Bolton P, Mira M, Kennedy P, Lahra MM. The quality of communication between hospitals and general practitioners: an assessment. *J Qual Clin Pract.* 1998;18:241-7.
303. Anderson MA, Helms LB. Communication between continuing care organizations. *Res Nurs.Health* 1995;18:49-57.
304. Brassey J, Elwyn G, Price C, Kinnersley P. Just in time information for clinicians: a questionnaire evaluation of the ATTRACT project. *BMJ* 2001;322:529-30.
305. Allery LA, Owen PA, Robling MR. Why general practitioners and consultants change their clinical practice: a critical incident study. *BMJ* 1997;314:870-4.
306. Kunz R, Ollenschläger G, Raspe H, Jonitz G, Kolkmann FW, (Eds.). Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis. Köln: Deutscher Ärzteverlag, 2000.
307. Greenhalgh T, Hughes J, Humphrey C, Rogers S, Swinglehurst D, Martin P. A comparative case study of two models of a clinical informaticist service. *BMJ* 2002;324:524-9.
308. Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, MacFarlane PW *et al.* Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. *N.Engl.J Med* 1995; 333:1301-7.

Anhang

Anhang Kapitel 2: Evidenztabellen

Anhang Kapitel 3:

3.1.: Checkliste Methodische Qualität von Leitlinien (8/1999)

3.2.: Gescreente und eingeschlossene Hypertonie-Leitlinien

Anhang Kapitel 4:

4.1: Berliner Gegenstandskatalogs

4.2: Testinstrument „Berliner Fragebogen“ (2 Fragensets)

Anhang Kapitel 5:

5.1. Dokumentationsbogen für das Telefon-Interview

5.2: Evaluationsbogen für das Pilot-Projekt „EbM-Informationen-Service

Anhang zu Kapitel 2: Evidenztabellen

Studie	Forschungsfrage	Methodischer Ansatz	Literatursuche / Sampling frame	Kritische Studienbewertung	Ergebnisse (Anzahl der Studien)	Schlußfolgerung (Autoren)
Britton A., Health Technol Ass 1998	Erzeugen RCTs systematisch größere oder kleinere Effektgrößen als nicht-randomisierte Studien (d.h. klinische Studien mit parallelen Kontrollen; klassische Kohorten)	Vergleich der Effektgrößen von RCTs und Nicht-RCTs in einzelnen Studien bzw. Kohorten von Studien, in denen vergleichbare Interventionen in vergleichbaren Settings mit vergleichbaren Endpunkten durchgeführt wurden Deskriptive Analyse	Systematische Literatursuche in MEDLINE, Embase, Science Citation Index, Cochrane Library Einschluß von 14 Einzelstudien und 4 Meta-Analysen	Detaillierte Beschreibung der Studien Methodische Bewertung: Wurde in der Analyse bei Imbalancen in den Baseline-Variablen adjustiert?	Kein Hinweis auf einheitliche Über- oder Unterschätzung des Behandlungseffekts durch Nicht-RCTs. <ul style="list-style-type: none"> • Gleiche Effekte (6 Einzelstudien/ 1 Kohorte) • Größere Effekte in Nicht-RCTs (5 Einzelstudien / 2 Kohorten) • Größere Effekte in RCTs (2 Einzelstudien / 1 Kohorte) • Widersprüchliche Ergebnisse (1 Einzelstudie) 	<ul style="list-style-type: none"> • Kein schlüssiges oder einheitliches Muster • Art der Intervention spielte keine Rolle • Kein Hinweis, dass Nicht-RCTs systematisch größere Schätzungen von Behandlungseffekten verursachen
Benson K. N Engl J Med 2000	Vergleich der Effekte von Beob.-Studien und RCTs bei vergleichbaren Therapien	Kriterien für Beob.-Studien: <ul style="list-style-type: none"> • nicht-experimentelle Studie • Vergleich von 2 Behandlungsalternativen • Arztbezogene Intervention • Kontrollgruppe Einschluß von Einzelstudien und MA Deskriptive Analyse	Literatursuche in MEDLINE / Cochrane Library (1985 – 1998) Einschluß: 3 Übersichten und 18 Einzelvergleichen (53 Beob.-Studien; 83 RCTs)	Keine systematische Bewertung der Studienmethodik (punktuelle Hinweise, z.B. Nachbeobachtung) Keine Kerninformationen über die eingeschlossenen Studien	<ul style="list-style-type: none"> • Einzelstudien (7 kardiologische, 11 nicht-kardiologische Interventionen) • Punktschätzer der Beob.-Studien innerhalb der KI der RCTs (2 Ausnahmen). • In 3 Studien qualitative Unterschiede (d.h., \emptyset Effekt vs. Benefit / Harm), aber innerhalb der KI. • Übersichten (n=3): Hormonersatz und Knochendichte / Calcium-Antagonisten und Nierentransplantat-Überleben / Infektionsrate bei laparoskopischer vs. offener Appendektomie. Vergleichbare Gesamtschätzer mit engen KI in beiden Designs	<ul style="list-style-type: none"> • Bei Bewertung von Interventionen liefern RCTs und Nicht-RCTs oft vergleichbare Ergebnisse → • Der Unterschied zu früheren Studien erklärt sich durch verbesserte Methoden der Beob.-Studien. • Stärkerer Nutzen von klinischen Datenbanken gefordert.
Concato J N Engl J Med 2000	Vergleich der Effekte von Beob.-Studien (FKS / Kohorten mit gleichzeitigen Kontrollen) und RCTs bei gleichen Behandlungen	Identifikation von MAs aus RCTs, Beob.-Studien oder gemischten Designs Kriterien: Gleiche Interventionen Dichotome Endpunkte Angabe der KI Deskriptive Analyse	Literatursuche in MEDLINE, Current Contents und Hand Searching MA aus 5 großen Medical Journals von '91-'95 (Annals of Internal Medicine; BMJ; JAMA; Lancet; New Engl J Med) Einschluß von Studien aus 5 Themenkomplexen (6 Publikationen mit 55 RCTs und 44 Beob.-Studien)	Bewertung anhand von veröffentlichten Kriterien, Ergebnisse nicht präsentiert Keine Kerninformationen über die eingeschlossenen Studien	<p>Gesamteffekt (95% KI)</p> <p>BCG-Vakzine + Tbc (RCTs vs. FKS): RR_{sum}: 0.49; (0.34–0.70) vs. OR_{sum}: 0.50; (0.39–0.65)</p> <p>Mammographie + Tod dr. Mamma-Ca (RCTs vs. FKS): RR_{sum}: 0.79; (0.71–0.88) vs. OR_{sum}: 0.61; (0.49–0.77)</p> <p>Cholesterin-Spiegel + Unfall (RCTs vs. Koh): RR_{sum}: 1.42; (0.94–2.15) vs. HR_{sum}: 1.40; (1.14–1.66)</p> <p>Hochdruck-Behandlung + Schlaganfall (RCTs vs. Koh): RR_{sum}: 0.58; (0.0.5–0.67) vs. adj PE_{sum}: 0.62; (0.60–0.65)</p> <p>Hypertonie + KHK (RCTs vs. Koh): RR_{sum}: 0.88; (0.78–0.96) vs. adj PE_{sum}: 0.77; (0.75–0.80)</p>	Kein Zusammenhang zwischen Qualitätskriterien und Größe des Effekts Im Vergleich zu RCTs keine systematische Effektüberschätzung in hochwertigen Kohorten / FKS

Studie	Forschungsfrage	Methodischer Ansatz	Literatursuche / Sampling frame	Kritische Studienbewertung	Ergebnisse	Schlußfolgerung (Autoren)
Ioannidis JP, JAMA 2001	<p>Liefern RCTs und Beob.-Studien bei gleichen Interventionen gleiche Ergebnisse ?</p> <p>(Def. Beob.-Studien: retro-, prospektive Kohorten, FKS, nicht-random. Trials mit gleichzeitigen und historischen Kontrollen)</p> <p>Exploration von Variablen zur Erklärung möglicher Diskrepanzen zwischen RCTs und Beob.-Studien</p>	<p>MA aus RCTs und Beob.-Studien mit mind. 1 primären dichotomen Endpunkt</p> <p>Analyse Gesamtschätzer von RCTs und Beob.-Studien (Random - / Fixed effects – Modell)</p> <p>Übereinstimmung</p> <ul style="list-style-type: none"> • Korrelation zwischen Gesamt-OR_{RCTs} / Gesamt-OR_{Beob.-Studien} • Häufigkeit, daß Gesamt-OR_{Beob.-Stud.} > Gesamt-OR_{RCT}? • Unterschied zwischen RCTs und Beob.-Studien größer als durch Zufall erwartet ? • Sensitivitätsanalyse f. verschiedene Definitionen von Diskrepanz: <ul style="list-style-type: none"> - OR_{Beob.-Stud.} \geq / \leq OR_{RCT} • In OR_{Beob.-Stud.} 50% \geq / \leq In OR_{RCT} <p>Erklärung von Diskrepanzen dr. Design; Publikationsjahr; Alter; Ereignisrate; Nachbeobachtung</p> <p>Quantitative Analyse</p>	<ul style="list-style-type: none"> • Reviews der Cochrane Library • MEDLINE (MA mit RCTs + Beob.-Studien) • Persönliche Datenbank mit MA (1991-1997) • Referenzlisten (Kunz, Britton, Benson, Concato) • Eigene MAs <p>Einschluß 45 MAs mit 240 RCTs und 168 Beob.-Studien (71 prospektive, 40 retrospektive Kohorten; 25 FKSs; 29 Studien mit historischen Kontrollen; 3 sonstige Studien)</p>	<p>Keine systematische Bewertung der Studienmethodik der MA</p> <p>Keine Kerninformationen über die eingeschlossenen Studien</p>	<p>Heterogenität innerhalb der Studientypen: RCTs: 23% (9/39 Topics); Nicht-RCTs: 41%; (13/32 Topics); prospektive Kohorten: 40% (6/15)</p> <p>Korrelationskoeffizient: 0.75 (alle); 0.83 (ohne historische Kontrollen); $p < 0.01$</p> <p>Effektgrößen:</p> <ul style="list-style-type: none"> • Nicht-RCTs > RCTs: 25/45 Topics • Nicht-RCTs < RCTs: 14/45 Topics <p>Diskrepanzen RCTs / Nicht-RCTs in Abhängigkeit von Analyse: Random effect: 16% (7/45 topics); Fixed effect: 27% (12/45 topics)</p> <p>Unterschied in ln OR: alle Designs: < 50%: 28/45 Fällen (62%) 2fache OR: 15/45 Fällen (33%)</p> <p>Vergleich prospektiver Studien: 50%: 16/26 Topics (62%) 2 fache OR: 7/26 Fälle (27%)</p> <p><u>Disagreement beyond chance</u> Günstigere Effekte sign. häufiger in Beob.-Studien als in RCTs ($p=0.03$ random effects; 0.02 fixed effects)</p> <p>Keine weiteren Erkenntnisse durch Analysen bzgl. Alter, Follow-up etc.</p>	<p>Deutliche Divergenz bzgl. der Effektgröße trotz guter Korrelation zwischen Effekten von RCTs and Beob.-Studien:</p> <ul style="list-style-type: none"> • Trend für Effekt-Überschätzung durch Beob.-Studien • nicht allgemeingültig, da auch gegenteilige Beobachtungen <p>Potentielle Ursachen für die Variabilität innerhalb und zwischen Designs:</p> <ul style="list-style-type: none"> • intrinsische Design-Unterschiede • Populationen (divergente Einschlußkriterien; natürliche Variabilität bei gleichen Kriterien) • kleines Sample • Zielsetzung: Efficacy vs. Effectiveness • Method. Mängel (in beiden Designs möglich) • Bias

Studie	Forschungsfrage	Methodischer Ansatz	Literatursuche / Sampling frame	Kritische Studienbewertung	Ergebnisse	Schlußfolgerung (Autoren)
Kunz R. Cochrane Database of Methodology Reviews Cochrane Library 2001	Wie ist die Auswirkung von Randomisierung / Nicht-Randomisierung auf Effektschätzer auf der Basis empirischer Daten? 4 Vergleiche: • RCT vs. Nicht-RCTs bei gleicher Intervention • RCT vs. Nicht-RCTs bei heterogenen Interventionen • Randomisierung mit maskierter vs. nicht-maskierter Verblindung • Qualitativ hochwertige vs. weniger hochwertige Studien explorativer Ansatz	Vergleich des Gesamteffekts von systematischen Übersichten, Meta-Analysen, Kohorten von Studien, in denen eine Intervention untersucht wurde. Sensitivitätsanalysen • ohne historische Kontrollen • ausschließlich methodisch hochwertige Studien Ausschluß von klassischen Beob.-Studien wie FKS oder Kohortenstudien Deskriptive Analyse	• Cochrane Review Methodology Database • Methodologische Bibliographien • MEDLINE • Science Citation Index • Handsearching • Persönliche Kommunikation mit Methodikern • Referenzlisten von relevanten Artikeln Gesamtbewertung von 32 systematischen Übersichten / Meta-Analysen oder Kohorten (> 3000 Einzelstudien)	Methodische Bewertung von 6 a priori festgelegten methodischen Items Qualitative Einteilung in: - Ohne wesentlichen methodischen Fehler - potentiell wichtiger methodischen Fehler - schwerwiegender methodischer Fehler Tabellarische Präsentation der Studien	Quantitative Zusammenfassung der Ergebnisse (Bezugsgröße ist RCT, angemessene Randomisierung oder hochwertige Studie) <u>1. Frage (n=15 Übersichten mit 35 Vergleichen)</u> • Effekt-Überschätzung dr. Nicht-RCTs (n=22) • Effekt-Unterschätzung dr. Nicht-RCTs (n=4) • Gleicher Effekt (n=8) • Umkehrung des Effekts in Nicht-RCTs (n=1) <u>2. Frage (n=7)</u> • Keine schlüssige Bewertung (n=3) • Keine konstante Beziehung zwischen Design/Qualität und Effekt (n=4) <u>3. Frage (n=3):</u> • Überschätzung dr. nicht angemessen maskierte Randomisierung (n=3) <u>4. Frage (n=9 Übersichten mit 11 Vergleichen):</u> • Überschätzung in weniger hochwertigen Studien (n=5) • Unterschätzung in weniger hochwertigen Studien (n=2) • Umkehrung des Effekts (n=2) • kein schlüssiger Zusammenhang (n=2) <u>Sensitivitätsanalyse:</u> • ohne historische Kontrollen (1. Frage): Ergebnisse vergleichbar zur Hauptfrage (p=0.7) s.o. • Hochwertige Studien: Trend zur Effekt-Überschätzung bei den verschiedenen Fragen, aber nicht ausschließlich	Hauptergebnis: Die Effekte mit (angemessener) Randomisierung sind häufig, aber nicht ohne Ausnahme kleiner; Verzerrungen können sich in alle Richtungen bewegen. Die Sensitivitätsanalyse ohne historische Kontrollen bestätigte das Hauptergebnis und widerlegt die von Concato geäußerte Hypothese, daß die 1998 im BMJ beschriebene Überschätzung auf Designs mit historischen Kontrollen zurückzuführen sind. Die Bestätigung der Hauptergebnisse bei ausschließlicher Berücksichtigung hochwertiger Studien unterstreicht die Aussagekraft der Studie

Checkliste Methodische Qualität von Leitlinien

2. Version (8 / 1999)

Gültig: bis 31. August 2002

Bewertungsinstrument des Leitlinien-Clearingverfahrens

von Bundesärztekammer und
Kassenärztlicher Bundesvereinigung
in Kooperation mit der
Deutschen Krankenhausgesellschaft
und den Spitzenverbänden der
Gesetzlichen Krankenkassen

Zentralstelle
der Deutschen
Ärzeschaft zur
Qualitätssicherung
in der Medizin

Gemeinsame
Einrichtung der
Bundesärztekammer
und der
Kassenärztlichen
Bundesvereinigung

Impressum

Herausgeber:

Zentralstelle der Deutschen
Ärzeschaft zur Qualitätssicherung
in der Medizin

Gemeinsame Einrichtung der
Bundesärztekammer und der
Kassenärztlichen Bundesvereinigung

Anschrift:

Ärztliche Zentralstelle Qualitätssicherung
Aachener Straße 233 – 237
D 50931 Köln

Telefon (0221) 4004-500/501

Telefax (0221) 4004-590

e-mail azq@dgn.de

Internet:

<http://www.azq.de>

<http://www.leitlinien.de>

2.Version

Fassung 8 / 1999

© 1999 Ärztliche Zentralstelle Qualitätssicherung

INHALT

EINFÜHRUNG	3
1. FRAGEN ZUR QUALITÄT DER LEITLINIENENTWICKLUNG.....	5
2. FRAGEN ZU INHALT UND FORMAT DER LEITLINIE	6
3. FRAGEN ZUR ANWENDBARKEIT DER LEITLINIE.....	6

Einführung

Bundesärztekammer, Kassenärztliche Bundesvereinigung, Deutsche Krankenhausgesellschaft und Spitzenverbände der Gesetzlichen Krankenversicherungen (BÄK, KBV, DKG, GKV) fördern und unterstützen gemeinsam wissenschaftlich begründete und praktikable Leitlinienprogramme in der Medizin. Sie sehen Leitlinien an als

- Hilfen für ärztliche Entscheidungsprozesse im Rahmen einer leistungsfähigen Versorgung der Patienten;
- wesentliche Bestandteile von Qualitätssicherung und Qualitätsmanagement;
- Instrumente zur Verbesserung der Versorgungsergebnisse, zur Minimierung von Behandlungsrisiken und zur Erhöhung der Wirtschaftlichkeit;
- Hilfen für die ärztliche Aus-, Weiter- und Fortbildung (1,2).

Zur Unterstützung wissenschaftlich begründeter und praktikabler Leitlinien, haben die Bundesärztekammer und Kassenärztlicher Bundesvereinigung 1997 ein Programm zur Qualitätsförderung von Leitlinien gestartet.

Mit Unterstützung namhafter Experten wurden folgende Instrumente und Maßnahmen entwickelt:

- Die **„Beurteilungskriterien für Leitlinien in der medizinischen Versorgung“** (Leitlinie für Leitlinien) beschreibt die Ziele und Qualitätsanforderungen für Leitlinien (1).
- Mit der **„Checkliste zur methodischen Qualität von Leitlinien“** lassen sich die Qualitätskriterien guter Leitlinien systematisch erfassen und dokumentieren (3,4) . Zur Checkliste ist ein **Leitlinien-Manual** erhältlich (5).
- Durch das **„Leitlinien-Clearingverfahren“** von BÄK und KBV in Kooperation mit DKG und GKV werden Leitlinien, die den Qualitätskriterien entsprechen, gekennzeichnet, gefördert (2) und über
- **Leitlinien-In-Fo“** – dem Online- Informations- und Fortbildungsprogramm der ÄZQ – bekannt gemacht (4).

Die **Checkliste zur methodischen Qualität von Leitlinien, die hiermit in einer 2., überarbeiteten Version, vorgelegt wird**, orientiert sich an ähnlichen Instrumenten aus den USA, Schottland und England (6-8).

Sie wurde durch Mitglieder des "Expertenkreises Leitlinien" der Ärztlichen Zentralstelle Qualitätssicherung erarbeitet, in Kooperation mit der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF).

Im Vergleich zur 1. Version (1998) wurden durch die Überarbeitung terminologische Unklarheiten beseitigt und Anregungen aus einem europäischen Projekt zur Leitlinienbewertung (9) aufgenommen.

Form der Checkliste und Anzahl der Fragen wurden beibehalten. Die Lektüre des Leitlinien-Manuals (5) wird weiterhin empfohlen.

Wie die Erfahrungen im In- und Ausland (8, 10) zeigen, kann eine solche "Checkliste" insbesondere für Autoren und Herausgeber, aber auch für Nutzer von Leitlinien hilfreich sein. Es wäre wünschenswert, wenn die vorliegende deutsche Version die zahlreichen Bemühungen um gute Leitlinien in der Medizin unterstützen könnte.

Die Checkliste zur methodischen Qualität von Leitlinien ist Bewertungsinstrument des Leitlinien-Clearingverfahrens (2).

Literatur:

1. Bundesärztekammer, Kassenärztliche Bundesvereinigung. Beurteilungskriterien für Leitlinien in der medizinischen Versorgung. Dtsch Ärztebl 94 (1997) A2154-2155, B-1622-1623, C-1754-1755
2. Bundesärztekammer, Kassenärztliche Bundesvereinigung. Das Leitlinien-Clearingverfahren in Zusammenarbeit mit der Deutschen Krankenhausgesellschaft und den Spitzenverbänden der Gesetzlichen Krankenversicherungen. Dtsch Ärztebl 94 (1999) A2105-2106. <http://www.leitlinien.de>
3. Ollenschläger G, Helou A, Kostovic-Cilic L, Perleth M, Raspe HH, Rienhoff O, Selbmann HK, Oesingmann U. Die Checkliste zur methodischen Qualität von Leitlinien. Z ärzt Fortb Qual sich (ZaeFQ) 92 (1998) 191-194
4. Ärztliche Zentralstelle Qualitätssicherung (Hrsg) Leitlinien-In-Fo. ÄZQ-Schriftenreihe Band 1. München, Zuckschwerdt 1999, S. 39-61. <http://www.leitlinien.de>
5. Helou A, Kostovic-Cilic L, Ollenschläger G. Nutzermanual zur Checkliste "Methodische Qualität von Leitlinien". Köln, Ärztliche Zentralstelle Qualitätssicherung. In (4)
6. Lohr KN, Field MJ. A provisional instrument for assessing clinical practice guidelines. In: Field MJ, Lohr KN (Hrsg) Guidelines for clinical practice. From development to use. Washington DC. National Academy Press 1992
7. Scottish Intercollegiate Guidelines Network. Clinical Guidelines - Criteria for Appraisal for National Use. <http://pc47.cee.hw.ac.uk/sign/critmain.html>
8. Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Moran S. Development and application of a generic methodology to assess the quality of clinical guidelines. Int Journ Qual Health Care 11 (1999) 21-28
9. Appraisal of Guidelines, Research & Evaluation in Europe (AGREE) Instrument. Draft, May 1999. BIOMED project PL 96-3669. Deutscher Kooperationspartner: Ärztliche Zentralstelle Qualitätssicherung.
10. Gerlach FM, Beyer M, Berndt M, Szecsenyi J, Abholz HH, Fischer GC. Das DEGAM-Konzept – Entwicklung, Verbreitung, Implementierung und Evaluation von Leitlinien für die hausärztliche Praxis. . Z ärzt Fortb Qual sich (ZaeFQ) 93 (1999) 111-120

1. Fragen zur Qualität der Leitlinienentwicklung

lfd.Nr	Verantwortlichkeit für die Leitlinienentwicklung	j	n	uk	na
1.1	Wird die für die Leitlinienentwicklung verantwortliche Institution klar genannt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
1.2.	Existieren detaillierte Angaben über finanzielle oder andere Formen der Unterstützung durch Dritte ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
1.3.	Falls Unterstützung seitens kommerzieller Interessengruppen erfolgte bzw. Hinweise auf mögliche Verpflichtungen / Interessenkonflikte existieren, wurde die mögliche Einflußnahme auf die Leitlinie diskutiert ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Autoren der Leitlinie	j	n	uk	na
1.4.	Sind die an der Erstellung der Leitlinie Beteiligten (Fach-, Interessen-, Patientengruppen) hinsichtlich ihrer Funktion und der Art ihrer Beteiligung klar genannt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
1.5.	Waren an der Erstellung der Leitlinie die von den Empfehlungen im wesentlichen Betroffenen (die mit der Thematik befaßten Fachdisziplinen und Patienten) beteiligt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Identifizierung und Interpretation der Evidenz	j	n	uk	na
1.6.	Werden Quellen und Methoden beschrieben, mit deren Hilfe die den Empfehlungen zugrunde liegenden Belege (Evidenz) gesucht, identifiziert und ausgewählt wurden ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
1.7.	Sind die Informationsquellen und Suchstrategien nachvollziehbar dokumentiert ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.8.	Werden die Methoden zur Interpretation und Bewertung der Evidenzstärke genannt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-

	Formulierung der Leitlinienempfehlungen	j	n	uk	na
1.9.	Werden die zur Auswahl der Leitlinienempfehlungen eingesetzten Methoden genannt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
1.10.	Werden die zur Konsentierung der Leitlinienempfehlungen eingesetzten Verfahren genannt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
1.11.	Wird erwähnt, wie die Vorstellungen von interessierten Gruppen, die nicht an der Leitlinienerstellung beteiligt waren, berücksichtigt wurden ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
1.12.	Ist die Verknüpfung der wichtigsten Empfehlungen mit der zugrunde liegenden Evidenz exakt dokumentiert ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Gutachterverfahren und Pilotstudien	j	n	uk	na
1.13.	Ist die Leitlinie vor der Veröffentlichung durch unabhängige Dritte begutachtet worden ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
1.14.	Werden die Methoden, Kommentierungen, Konsequenzen einer Begutachtung erwähnt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.15.	Wurde die Leitlinie einer Vortestung oder einem Pilotversuch unterzogen ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
1.16.	Werden die Methoden, Ergebnisse und Konsequenzen eines Pilotversuchs erwähnt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.17.	Wurde die Leitlinie mit anderen thematisch vergleichbaren Leitlinien verglichen ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Gültigkeitsdauer / Aktualisierung der Leitlinie	j	n	uk	na
1.18.	Ist in der Leitlinie ein Zeitpunkt genannt, zu dem sie überprüft / aktualisiert werden soll ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.19.	Sind Zuständigkeit und Verfahrensweisen für Überprüfung / Aktualisierung klar definiert ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Transparenz der Leitlinienerstellung	j	n	uk	na
1.20.	Wurden die möglichen systematischen Fehler / Konflikte umfassend diskutiert ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
1.21.	Existiert eine zusammenfassende Darstellung über den Inhalt, die Empfehlungen der Leitlinie sowie über die Methodik der Erstellung (z.B. in Form eines Leitlinien-Reports) ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-

Nur eine Antwortkategorie ankreuzen:

j: ja

n: nein

uk: unklar (auch bei unvollständigen / unzureichenden Angaben)

na: nicht anwendbar

2. Fragen zu Inhalt und Format der Leitlinie

	Ziele der Leitlinie	j	n	uk	na
2.1.	Sind die Gründe für die Leitlinienentwicklung explizit genannt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
2.2.	Sind die Ziele der Leitlinie eindeutig definiert ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-

	Kontext (Anwendbarkeit / Flexibilität)	j	n	uk	na
2.3.	Ist (sind) die Patienten-Zielgruppe(n) der Leitlinie eindeutig definiert (z.B. hinsichtlich Geschlecht, Alter, Krankheitsstadium, Begleiterkrankungen usw.) ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.4.	Sind die Anwender, an die sich die Leitlinie richtet, eindeutig definiert (z.B. die ärztliche Zielgruppe) ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.5.	Enthält die Leitlinie Angaben über Situationen, in denen spezielle Empfehlungen der Leitlinie nicht berücksichtigt werden können oder sollen ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.6.	Wurden Ansichten, Präferenzen und mögliche Reaktionen der Patienten-Zielgruppe(n) berücksichtigt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Klarheit, Eindeutigkeit	j	n	uk	na
2.7.	Beschreibt die Leitlinie das behandelte Gesundheits- / Versorgungsproblem eindeutig und in allgemein verständlicher Sprache ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
2.8.	a) Sind die Empfehlungen logisch, inhaltlich konsistent, eindeutig, leicht nachvollziehbar dargestellt und übersichtlich präsentiert ? b) Enthält die Leitlinie wesentliche (Schlüssel-) Empfehlungen, die leicht zu identifizieren sind ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
2.9.	Sind die in Frage kommenden Handlungsalternativen und die Entscheidungskriterien für ihre Auswahl präzise beschrieben ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
2.10.	Liegen differenzierte Empfehlungen zur Entscheidung hinsichtlich ambulanter oder stationärer Versorgung vor ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-
2.11.	Existieren Angaben darüber, welche Maßnahmen notwendig erscheinen ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.12.	Existieren Angaben über Maßnahmen, die unzureichend, überflüssig, obsolet erscheinen ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Nutzen, Nebenwirkungen, Kosten, Ergebnisse	j	n	uk	na
2.13.	Wird der bei Befolgen der Leitlinie zu erwartende gesundheitliche Nutzen bezüglich z.B. Morbidität, Mortalität, Symptomatik, Lebensqualität genannt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.14.	Werden die bei Befolgen der Leitlinie möglichen Risiken (Nebenwirkungen und Komplikationen) der Diagnostik / Therapie genannt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.15.	Wurden bei der Formulierung der Empfehlungen die Folgen für Kosten und andere Ressourcen berücksichtigt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.16.	Wird eine Abwägung der möglichen Vorteile, Risiken, Kosten vorgenommen und unterstützt dies die vorgeschlagene Vorgehensweise ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Nur eine Antwortkategorie ankreuzen:

- j:** ja
n: nein
uk: unklar (auch bei unvollständigen / unzureichenden Angaben)
na: nicht anwendbar

3. Fragen zur Anwendbarkeit der Leitlinie

Verbreitung und Implementierung		j	n	uk	na
3.1.	a) Existieren Instrumente / Maßnahmen, die die Anwendung der Leitlinie unterstützen können ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b) Wurden die möglichen Probleme bezüglich der Einstellungs- oder Verhaltensänderungen von Ärzten und anderen Leistungserbringern im Gesundheitswesen bei Anwendung der Leitlinie berücksichtigt ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	c) Wurden die möglichen organisatorischen Hindernisse der Leitlinien-Anwendung berücksichtigt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.2.	Nur für überregionale Leitlinien: Existieren Empfehlungen zur Methode der regionalen Anpassung der Leitlinie ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Überprüfung der Anwendung		j	n	uk	na
3.3.	Wird in der Leitlinie erwähnt, wie aus den Empfehlungen meßbare Kriterien / Indikatoren abgeleitet werden können, um das Befolgen der Leitlinie zu ermitteln ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.4.	Werden meßbare Kriterien / Indikatoren genannt, anhand derer der Effekt der Leitlinienanwendung überprüft werden kann ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Nur eine Antwortkategorie ankreuzen:

- j:** ja
n: nein
uk: unklar (auch bei unvollständigen / unzureichenden Angaben)
na: nicht anwendbar

Leitlinien/Literatur zum Leitlinien-Clearing-Bericht Hypertonie

Bewertete Leitlinien

Nicht bewertete Leitlinien

Dubletten

Recherchierte Literatur

Im Leitlinienbericht bewertete Hypertonieleitlinien:

1. Arzneimittelkommission der Deutschen Ärzteschaft: Empfehlungen zur Therapie der arteriellen Hypertonie, Köln, 1998 Arzneiverordnung in der Praxis. Sonderheft 9 (1998) ISSN 0939-2017 (Lang- und Kurzfassung. Z ärztl Fortb Qual sich 92 (1998) 701-704 (Kurzfassung). Bluthochdruck – Eine Information für Patienten und Angehörige. Techniker Kasse Hamburg (1. Aufl. 1999) ISBN 3-933779-02-2 (Patientenversion). Leitlinien- Report: Lasek R, Müller-Oerlinghausen B (1997) Therapieempfehlungen der Arzneimittelkommission der deutschen Ärzteschaft. Z ärztl Fortb Qual sich 91: 375-383. <http://www.akdae.de>
2. British Hypertension Society: Guidelines for management of hypertension: report of the 3rd working party of the British Hypertension Society. Journal of Human Hypertension 13(1999) 569-592, British Medical Journal 319 (1999) 630-635. <http://www.hyp.ac.uk/bhs/managemt.html>, <http://www.bmj.com/cgi/content/full/319/7210/630>
3. Canadian Coalition for High Blood Pressure Prevention and Control: Lifestyle modifications to prevent and control hypertension. Canadian Medical Association Journal 1999; 160 (9 Suppl): S1-S50. <http://www.cma.ca/cmaj/vol-160/issue-9/hypertension/index.htm>
4. Chalmers J et al. WHO-ISH Hypertension Guidelines Committee. 1999 Guidelines for Management of Hypertension. Journal Hypertension 1999, 17:151-185. http://www.who.int/ncd/cvd/HT_guide.html
5. Department of Veterans Affairs: The Pharmacologic Management of Hypertension. US Veterans Health Administration Publ. No. 96-0003, 12 / 1996. Aktualisierte Version (Hypertension Guideline 2000), 24 S.: <http://www.va.gov/health/hypertension/HTN.doc>
6. Deutsche Liga zur Bekämpfung des hohen Blutdruckes e.V., Deutsche Hypertonie Gesellschaft: Arterielle Hypertonie, 1998, 16 S., <http://www.uni-duesseldorf.de/WWW/AWMF/II/ihypto01.htm>
7. Feldman R, Campbell N, Larochelle P, for the Task Force for the Development of the 1999 Canadian Recommendations for the Management of Hypertension . 1999 Canadian recommendations for the management of hypertension . CMAJ 1999;161(12 Suppl):S1 ff <http://www.cma.ca/cmaj/vol-161/issue-12/hypertension/index.htm>
8. Hypertension Society of South Africa (1995) Guidelines for the management of hypertension at primary health care level. S Afr Med J 1995; 95: 1321-1325 (Leitlinien-Report:1325-1338)
9. Institute for Clinical Systems Integration: Hypertension _ Diagnosis and treatment, Bloomington (MN): 1999. Aktualisierte Version (2000), 41 S.: <http://www.icsi.org/guide/HTN.pdf> .
10. National High Blood Pressure Education Program. Sixth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. 1997, 70 S., NIH Publication No. 98-4080. <http://www.nhlbi.nih.gov/guidelines/hypertension/jncintro.htm>

11. New Zealand Guidelines Group: Guidelines for the management of mildly raised blood pressure in New Zealand. National Health Committee, PO Box 5013, Wellington.
http://www.nzgg.org.nz/library/gl_complete/bloodpressure/

Nicht bewertete Leitlinien:

1. Abbott D, et al. Guidelines for measurement of blood pressure, follow-up, and lifestyle counselling. Canadian Coalition for High Blood Pressure Prevention and Control. Can J Public Health. 1994, 85 Suppl 2:S29-243.
2. ANAES: Diagnostic et traitement de l'hypertension artérielle essentielle de l'adulte de 20 à 80 ans. 1997 In: Service des Références Médicales, S. 241-277
3. Canadian Hypertension Society: Report of the Canadian Hypertension Society Consensus Conference, 1992. Part 1: Introduction. CMAJ 149(3):289-293, Part 2: Diagnosis of hypertension in adults. CMAJ 149(4):409-418, Part 3: Pharmacologic treatment of essential hypertension. CMAJ 149(5): 575-584, Part 4: Hypertension in the elderly. CMAJ Vol. 149(6): 815-820, Part 5: Hypertension and diabetes. CMAJ 149(6): 821-826
4. Canadian Hypertension Society: Report of the Canadian Hypertension Society Consensus Conference 1997. [Part 1: Definitions, evaluation and classification of hypertensive disorders in pregnancy. CMAJ 1997; 157:715-725](#), [Part 2: Nonpharmacologic management and prevention of hypertensive disorders in pregnancy. CMAJ 1997, 157:907-919](#), [Part 3: Pharmacologic treatment of hypertensive disorders in pregnancy. CMAJ 1997, 157:1245-1254](#)
5. CBO: Richtlijnen Hypertensie, 1990.
6. Joint National Committee...: The fifth report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure...(JNC-V). Arch Intern Med 1993, 153:154-183.
7. Logan AG: Screening for hypertension in young and middle-aged adults. In: Canadian Guide to Clinical Preventive Health Care. Ottawa: Health Canada, 1994, 636-648
8. National Heart, Lung and Blood Institute: National High Blood Pressure Education Program Working Group report on primary prevention of hypertension, 1993, 53 S.
<http://www.nhlbi.nih.gov/health/prof/heart/hbp/pphbp.htm>
9. National Heart, Lung and Blood Institute: National High Blood Pressure Education Program Working Group: Update on the Task Force Report (1987) on high blood pressure in children and adolescents, 1996, VI, 24 S.,
http://www.nhlbi.nih.gov/health/prof/heart/hbp/hbp_ped.htm
10. National Heart, Lung and Blood Institute: National High Blood Pressure Education Program Working Group report on hypertension in diabetes, 1994, 33 S.,
<http://www.nhlbi.nih.gov/health/prof/heart/hbp/hbpd diab.txt>
11. National Heart, Lung and Blood Institute: National High Blood Pressure Education Program Working Group report on high blood pressure in pregnancy
12. National Heart, Lung and Blood Institute: National Heart, Lung and Blood Institute: National High Blood Pressure Education Program Working Group report on hypertension in the elderly. Hypertension 1994, 23(3): 275-285
13. Nederlands Huisartsen Genootschap: NHG-Standaard: Hypertensie 1991. In: NHG-Standaarden/Nederlands Huisartsen Genootschap
14. Patterson C, Logan AG: Hypertension in the elderly: case-finding and treatment to prevent vascular disease. In: Canadian Guide to Clinical Preventive Health Care. Ottawa: Health Canada, 1994, 944-951
15. Screening for hypertension. Guide to clinical preventive services. 2nd ed. Baltimore: Williams & Wilkins, 1996, S. 39-51,
<http://odphp.osophs.dhhs.gov/pubs/guidecps/text/CH03.txt>

16. World Health Organization/International Society: 1993 guidelines for the management of mild hypertension: memorandum from a World Health Organization/International Society of Hypertension meeting. Guidelines Sub-Committee. J Hypertens. 1993, 11(9):905-918.

Dubletten:

1. Joint National Committee...: Sixth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Arch Intern Med 1997; 157: 2413-2446
2. ANAES: Diagnostic et traitement de l'hypertension artérielle essentielle de l'adulte de 20 à 80 ans. Journal des Maladies Vasculaires 1998, 23 (3): 204-231
3. World Health Organization-International Society of Hypertension: IX-1993 guidelines for management of mild hypertension. WHO/ISH Mild Hypertension Liaison Committee. Clin Exp Hypertens. 1993, 15(6):1363-1395.
4. Patterson C: Hypertension in the elderly: case-finding and treatment to prevent vascular disease.
5. Screening for hypertension in young and middle-aged adults

Recherchierte Literatur zu Hypertonieleitlinien:

1. [No authors listed] 1993 guidelines for the management of mild hypertension: memorandum from a WHO/ISH meeting. Bull World Health Organ. 1993;71(5):503-17.
2. [No authors listed] American Society of Hypertension releases guidelines on home and ambulatory blood pressure monitoring. Am Fam Physician. 1996 Sep 15;54(4):1390.
3. [No authors listed] Guidelines for the management of hypertension at primary health care level. Hypertension Society of Southern Africa, endorsed by the Medical Association of South Africa and the Medical Research Council. S Afr Med J. 1995; 85(12 Pt 2):1321-5.
4. [No authors listed] IX-1993 guidelines for management of mild hypertension. WHO/ISH Mild Hypertension Liaison Committee. Clin Exp Hypertens. 1993 Nov;15(6):1363-95.
5. [No authors listed] Management guidelines for hypertension in southern Africa. S Afr Med J. 1992; Suppl:1-4. Review.
6. [No authors listed] The fifth report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure. Arch Intern Med. 1993;153(2):154-83.
7. [No authors listed] The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. Arch Intern Med. 1997; 24;157(21):2413-46.
8. 1991 guidelines for the prevention of hypertension and associated cardiovascular disease. Joint World Health Organization/International Society of Hypertension Meeting. J Hypertens Suppl. 1991; 9 (8): S60-2
9. Adams JL, et al.: A novel method of guideline development for the diagnosis and management of mild to moderate hypertension. Br J Gen Pract 1999; 49(440): 175-9
10. Alexander LM: Guidelines for hypertension treatment: applications for primary care practice – a review of the JNC VI report. Lippincotts Prim Care Pract 1998 Sep-Oct; 2(5):485-97
11. Becker A: Pharmakotherapie der Hypertonie in der hausärztlichen Praxis. Z. Allg. Med. 1999, 75:613-16
12. [British Cardiac Society, British Hyperlipidaemia Association, British Hypertension Society, British Diabetic Association. Joint British recommendations on prevention of coronary heart disease in clinical practice: summary. BMJ 2000;320:705-708.](#)

13. Chalmers J. Round table 2. First-line drugs: the position of the World Health Organization-International Society of Hypertension. J Hypertens Suppl. 1993;11 Suppl 5:S381-3.
14. Chalmers J: Impact of guidelines for antihypertensive treatment on clinical practice. Clin Exp Hypertens 1996; 18(3-4):415-23
15. Chalmers J: Implementation of guidelines for management of hypertension. Clin Exp Hypertens 1999; 21(5-6): 647-57
16. Clem JR: Recent changes in recommendations for antihypertensive therapy. Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure. S D J Med. 1993; 46(4):122-3.
17. Deutsche Liga zur Bekämpfung des hohen Blutdruckes e.V. Deutsche Hypertonie-Gesellschaft: Hypertonie-Therapie: Neueste Empfehlungen, Teil 1. Münch. med. Wschr. 1996, 138(51/52): 856-860.
18. Deutsche Liga zur Bekämpfung des hohen Blutdruckes e.V. Deutsche Hypertonie-Gesellschaft: Hypertonie-Therapie: Neueste Empfehlungen, Teil 3. Münch. med. Wschr. 1997, 139(1/2): 18-20.
19. Dickerson JE, Garratt CJ, Brown MJ: Management of hypertension in general practice: agreements with and variations from the British Hypertension Society guidelines. J Hum Hypertens 1995 Oct; 9 (10): 835-9
20. Donner-Banzhoff N: "Diagnose Hypertonie" – warum gibt es internationale Unterschiede? Z. Allg. Med. 1999, 75:604-12
21. Frohlich ED: Impact and interpretation of multiple guidelines and consensus documents. J Hum Hypertens 1996; 10 Suppl. 1: S63-7
22. Furberg CD, et al.: Overtreatment and undertreatment of hypertension. J Intern Med 1994; 235(5):387-97
23. Gilbert RE, et al. Diabetes and hypertension. Australian Diabetes Society position statement. Med J Aust. 1995;163(7):372-5.
24. Hetlevik I, et al: Implementing clinical guidelines in the treatment of hypertension in general practice. Blood Press 1998; 7(5-6): 270-6
25. Hirschl MM, et al.: Guidelines for the drug treatment of hypertensive crises. Drugs 1995; 50(6): 991-1000
26. Hoffbrand B, Ross M: Hypertension. In: Clinical Guidelines. London: The Royal College of General Practitioners, 1992.
27. International Society of Hypertension. J Hypertens Suppl. 1995;13(2):S119-22. Review.
28. [Jackson R. Guidelines on preventing cardiovascular disease in clinical practice. Absolute risk rules but raises the question of population screening. BMJ 2000;320:659-661.](#)
29. [Jackson R. Updated New Zealand cardiovascular disease risk-benefit prediction guide. BMJ 2000;320:709-710.](#)
30. Kaplan NM. Guidelines for the treatment of hypertension: an American view. Fifth Joint National Committee. J Hypertens Suppl. 1995; 13(2):S113-7. Review.
31. Magometschnigg D, et al [Individualized therapy of hypertension. The Austrian Professional Group of Clinical Pharmacology]. Wien Med Wochenschr. 1995;145(15-16):349-52.
32. Meredith PA. New FDA guidelines on the treatment of hypertension: comparison of different therapeutic classes according to trough/peak blood pressure responses. Arch Mal Coeur Vaiss. 1994; 87(11):1423-9.
33. Meyer-Sabellek WA: Proposals for European Commission guidelines in the use of ambulatory blood pressure monitoring to evaluate antihypertensive drugs. J of Hypertension 1991, 9 (Suppl. 8): S60-S62
34. Moser M: National recommendations for the pharmacological treatment of hypertension: should they be revised? Arch Intern Med 1999; 159(13): 1403-6

35. Myers MG: Ambulatory blood pressure monitoring: a proposal for new Canadian guidelines. Division of Cardiology, Toronto, Ontario. Can J Cardiol 1999; 15 Suppl. C:7C:12C
36. National Cholesterol Education Program. Second Report of the Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II). Circulation 1994; 89(3)
37. O'Connor PJ, et al.: Impact of hypertension guideline implementation on blood pressure control and drug use in primary care clinics. Jt Comm J Qual Improv 1999; 25(2): 68-77
38. Onion CW, et al.: Local clinical guidelines: description and evaluation of a participative method for their development and implementation. Fam Pract 1996 Feb; 13 (1): 28-34
39. [Pickering TG: Advances in the treatment of hypertension. JAMA1999; 281:114-116](#)
40. Ramsay LE et al.: The rationale for differing national recommendations for the treatment of hypertension. Am J Hypertens 1998; 11(6 Pt2): 79S-88S; discussion 95S-100S
41. Ramsey SD, et al.: An economic evaluation of the JNC hypertension guidelines using data from a randomized controlled trial. Joint National Committee. J Am Board Fam Pract 1999; 12(2):105-14
42. Rayburn WF Sixth report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure. A summary. J of Reproductive Medicine 1998; 43(5):444-50
43. [Robson J, Boomla K, Hart B, Feder G. Estimating cardiovascular risk for primary prevention: outstanding questions for primary care BMJ 2000;320:702-704.](#)
44. Rose VL. National Heart, Lung, and Blood Institute releases new guidelines for the treatment of hypertension. Am Fam Physician. 1998; 57(2):362-4, 366.
45. Schwartz GL, Sheps SG: A review of the sixth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Curr Opin Cardiol 1999 Mar; 14 (2):161-8
46. Sever P, et al. Management guidelines in essential hypertension: report of the second working party of the British Hypertension Society. BMJ 1993; 306(6883):983-7.
47. Simpson IJ, et al. Management guidelines for progressive chronic renal failure. New Zealand Nephrologists Consensus Group. N Z Med J. 1997; 110(1055):421-3. Review.
48. Statement on ambulatory blood pressure monitoring by the German Hypertension League. J Hum Hypertens. 1995; 9(9):777-9.
49. Vallee JP. Sixth report of the Joint National Committee on the prevention, detection, evaluation and treatment of arterial hypertension. Presse Med. 1999; 28(16):853-8. French.
50. WHO/ISH Mild Hypertension Liaison committee. BMJ 1993; 307(6918):1541-6.
51. Zanchetti A, et al. : The 1993 guidelines for the management of mild hypertension: memorandum from a WHO/ISH meeting. Blood Press.1993; 2(2):86-100.
52. Zanchetti A: Guidelines for the management of hypertension: the World Health Organization/International Society of Hypertension view. J of Hypertension 1995, 13 (Suppl. 2):S119-S122

Lernziele und Lerninhaltskatalog für den TTT-Kurs und die Basisseminare

Liebe Teilnehmer des TTT-Vorbereitungsseminars,

Nach dem gemeinsamen Brainstorming auf dem TTT-Vorbereitungstreff in Berlin haben wir die verschiedenen Vorschläge in den Entwurf des Lerninhaltskatalogs eingearbeitet, den Sie als Appendix des 2. Rundschreibens erhalten. Obwohl ich mir von dem Vorbereitungstreff erhofft hatte, wir könnten den bereits vollen Katalog kürzen, wurde er durch die durchaus nützlichen und berechtigten Ergänzungen noch länger. Zusammen mit den allgemeinen Lernzielen (Appendix 1) wird damit das Raster für die TTT-Seminare vorgegeben.

Die Intention des TTT-Kurses ist die kritische Anwendung von Studienergebnissen im (Mediziner- Gutachter-, Beratungs-) Alltag. Das bedeutet, wir wollen kein Magisterprogramm in klinischer Epidemiologie entwickeln, sondern Ärzten helfen, Studien kritisch zu lesen und zu verstehen. Dabei werden wir immer inkomplett bleiben und müssen Mut zur Lücke beweisen und diesen Mut auch von unseren zukünftigen Tutoren fordern. Da wir uns einig sind, dass es nicht möglich ist, alle Lerninhalte und den gesamten Lernzielkatalog (geschweige denn alle sonstigen wichtigen klinisch epidemiologischen Inhalte) durch die beiden 2tägigen Seminare abzudecken, müssen etliche Programmpunkte dem Selbststudium der Teilnehmer vorbehalten bleiben. Somit sollen die vorgegebenen Lerninhalte den Trainees als Stütze dienen, um den eigenen Kenntnisstand zu kontrollieren und zu überprüfen.

Zusätzlich dienen die Lernziele und Lerninhalte als Matrix für Basisveranstaltungen. Da diese in oft sehr unterschiedlichem Format und Settings stattfinden, ist eine zu rigide Festlegung eher hinderlich als hilfreich.

Wir bitten Sie, Lernziele und Lerninhalte nochmals kritisch zu überprüfen und uns Ihre Kommentare und Änderungsvorschläge innerhalb von 2 Wochen zurückzuschicken.

Mit freundlichen Grüßen

Regina Kunz

Norbert Donner-Banzhoff

Lerninhalte für Trainerkurse (Matrix für Basisveranstaltungen)

Einführung und Allgemeine Aspekte

- **Definition EBM**

Begriffsklärung: Klinische Epidemiologie vs. evidenzbasierte Medizin

- **Arten von Studien: Experimentell vs. Beobachtung**

Experimentelle Studien

- randomisierte kontrollierte Studie

Beobachtungsstudien

- Kohortenstudie
- Fall-Kontroll-Studie
- Querschnittsstudie, Registerstudie, Korrelationsstudie, ökologische Studie, Fallserie

Zuordnung von klinischen Fragestellungen zu geeigneten Studiendesigns

- **Ursachen von fehlerhaften Studienergebnissen**

Bias: Definition und häufige Formen

Confounding: Definition und Beispiele

Chance: Definition und Beispiele

- **Stichprobenziehen / Sampling**

- **Effektmaße** (Allgemeines; Details bei einzelnen Studienarten)

- **Reliabilität und Validität von Daten**

- **Outcome-Messung**

Surrogatparameter vs. klinisch relevante Endpunkte

- **Strukturiertes Vorgehen zu Problemlösung (nach Sackett)**

- Beantwortbare Frage formulieren, Literatursuche, Kritische Bewertung, Anwendung auf den Patienten

- **Levels of Evidence**

- **Ethische Aspekte**

Statistik

- **Grundlagen**

- Deskriptive vs. inferentielle Statistik
- Skalen-Niveau (Kategorische vs. stetige Daten)

- **Differenzierung Population / Stichprobe**

Grundsätzliches Ziel statistischer Tests

- Bedeutung der Signifikanz, p-Wert; Fehler 1. und 2. Art

Robustheitsmerkmale statistischer Angaben

- Signifikanzniveau, Konfidenzintervall

Fehlerquellen bei statistischen Verfahren

- Zufällige (e.g. Power) und systematische (Bias) Fehler
- Multiples Testen, Untergruppen

- **Prinzipien wichtiger statistischer Verfahren**

Uni- und bivariate Analysen

- t-test, Chi² Test, Exact-tests

Adjustierte Analysen

- Stratifizierte Analysen
- Multi-variate Analysen

Überlebensanalyse

- Kaplan-Meier-Analyse

- **Differenzierung statistische Signifikanz / klinische Bedeutung**

Therapie / Intervention

- **Studiendesign**

explanative Studie (efficacy) vs. Management-Studie (effectiveness)

- **Experimentelle Studien**

- Randomisierte kontrollierte Studie
Randomisierung per se,
Randomisierung Individuen / Cluster
Probleme von fehlender Randomisierung (parallele / historische Kontrolle)

- **Wichtige Biasformen für Interventionsstudien**

- Selektionsbias, Performancebias (Struktur-, Behandlungs-, Gleichheitsbias) Attrition (Verlust-) Bias

- **Methodische Manöver / Merkmale**

- Verblindete Zuordnung (concealed allocation)
- Blinding

- **Art der Ergebnisanalyse**

- Intention to treat / Explanativ (Valid-case Analyse)
- Untergruppenanalyse

- **Beschreibung des Therapie-Effektes**

- Relatives und absolutes Risiko
- Relative und absolute Risikoreduktion (bzw. -anstieg)
- Bedeutung des Ausgangsrisikos

Number needed to treat / to harm

Übertragung auf individuellen Patienten (Adaptationsmöglichkeiten, z.B. Sackett's Faktor „f“)

Diagnostik / Screening

- **Ziel von diagnostischen Tests**

Reflexion über „Medical decision making“

- **Gold-Standard**

Definition

Fehlen eines objektiven Standards

- **Auswahl der Studienpopulation**

- **Wichtige Bias-Formen für Studien zu diagnostischen Tests**

- Measurement bias, Patienten-Spektrum; Verifikations- oder Work-up-Bias;
- Lead time Bias; Length time Bias (Screening)

- **Kennzahlen diagnostischer Tests**

- Sensitivität, Spezifität
- Likelihood-Ratio
- Positiver / negativer prädiktiver Wert
- ROC-Kurven
- Bayes Theorem

Genauigkeit, Präzision, Accuracy

Klinische Umsetzung

- Vortest-Wahrscheinlichkeit, Nachtest-Wahrscheinlichkeit
- Fagan's Likelihood Nomogramm

- **Screeningtests**

- Unterschiede zwischen klinischen diagnostischen Tests und Screening
- Kriterien für den Einsatz von Screening
- Probleme von Screening (e.g. Langzeitnutzen)

Indikationen

Prognose

- **Natürlicher Krankheitsverlauf vs. klinischer Verlauf**
 - **Arten von Prognosestudien**
 - Kohortenstudie (zeitlicher Verlauf)
Inzeptions-, Überlebenskohorte
 - Prospektive vs. retrospektive Kohortenstudie
 - **Wichtige Bias-Formen für Kohortenstudien (Auswahl)**
 - Selektionsbias, Migrationsbias, Measurement Bias
 - **Prognostische Kennzahlen**
 - Prävalenz und Inzidenz
 - Morbidität und Mortalität
 - **Risikofaktoren vs. prognostische Faktoren**
 - **Maßzahlen zur Beschreibung prognostischer Wahrscheinlichkeiten**
 - Relatives Risiko, Absolutes Risiko, Odds-Ratio
 - Beziehung : OR \Leftrightarrow RR
 - **Häufige statistische Verfahren**
 - Überlebenskurve (Kaplan-Meier)
 - Cox Proportional Hazards Modell (Erklärung des Prinzips)
 - **Umsetzung prognostischen Informationen**
 - Zielgruppe: Patienten / Ärzte vs. Öffentlichkeit
- Kommunikation der Informationen

Systematische Übersicht / Meta-Analyse

- **Narrative vs. systematische Übersicht; Meta-Analyse**

- Systematische Übersicht versus Meta-Analyse
- Sonderformen (Quelle publizierte Studien vs. Original-Patientendaten)

- **Rationale, Indikation, und Kontraindikation für Meta-Analysen**

- Power, Subgruppen-Analyse, Erhöhte Präzision des Schätzwerts

- **Durchführung (Prinzip)**

- **Maßzahlen der Ergebnis-Zusammenfassung von Meta-Analysen**

- Odds Ratio, Relatives Risiko; Risikodifferenz;
- Konfidenz-Intervall
- Graphische Darstellung

- **Grenzen von Meta-Analysen**

Ursachen für falsch-positive Ergebnisse

- Publication-Bias
- Language Bias
- Methodische Qualität der Primärpublikationen
- Mega-Trials als „Goldstandard“ für Meta-Analysen
- Chance (wenige Studien mit kleinen Patientenzahlen)

- **Kontrollmechanismen (Validitätskriterien)**

- Homogenität und Heterogenität (visuell und statistisch)
- Sensitivitätsanalyse
- Funnelplot (Publikationsbias)

Bedeutung der Literatur-Recherche

Kausalität/ Nebenwirkung

- **Fall-Kontroll (FK)-Studien / Kohorten / Surveillance**

- Studiendesign der Wahl bei seltenen Ereignissen
- Abgrenzungen der einzelnen Studiendesigns; Stärken/Schwächen
- Zeitverlauf Fall-Kontroll vs. Kohorten vs. Surveillance
- Assoziation vs. Kausalität; Kriterien für Kausalität

- **Methodische Manöver**

- Identifikation von Fällen
- Wahl der Kontrollen
- Matching / Stratifizierung
- Erhöhtes Verhältnis Fall-Kontrollen

- **Confounding und mögliche Kontrollmanöver**

- **Wichtige Biasformen bei FK- / Kohorten/ Surveillance-Studien**

- z.B. Selektionsbias, Verordnungsbias (prescribing bias), Interviewer Bias, Überweisungsbias (referral bias)

- **Maßzahlen für die Beschreibung einer Assoziation**

- Odds Ratio
- Konfidenzintervall
- Uni- vs. multivariate Analysen
- Identifikation von Risikofaktoren

Planung / Durchführung / Auswertung der Studien

Leitlinien

- **Allgemeines:** Abgrenzung gegenüber Standards und Richtlinien
- **Qualitätskriterien von Leitlinien**
 - Angabe und Einstufung der zugrundegelegten Studien (Evidenz / Levels of Evidence/ Grade of Recommendation)
 - Transparenz des Erstellungsverfahrens: Angabe der Beteiligten; Beschreibung der zugrundegelegten Sichtweise: Allgemein- und Fachärzte, Patienten, Kostenträger, Gesellschaft,
 - Angabe Erstellungszeitraums / Gültigkeitsdauer
- **Klarheit**
 - Klare Problemstellung
 - Beschreibung der Patientenzielgruppe
 - Klare Zielstellung der Leitlinie
 - Verständliche und eindeutige Formulierung der Empfehlungen
- **Kosteneffekte der Leitlinienumsetzung**
- **Flexibilität**
 - Berücksichtigung absehbarer Ausnahmefälle
 - Berücksichtigung individueller Patientenwünsche
- **Multidisziplinarität**
 - Berücksichtigung der Ansätze möglichst aller beteiligten Fachgruppen
 - Akzeptables Ergebnis bei Zustimmung aller beteiligten Gruppen
- **Arten von Leitlinien**
 - Zentrale (Fachgremien / Fachgesellschaften) vs. dezentrale Erstellung
 - Potentielle Vor- und Nachteile
- **Aktuelle Quellen von Leitlinien in Deutschland und international**
 - Clearingstelle für Leitlinien, Schottische Richtlinien, Agency of Healthcare, Research and Quality (AHRQ), AWMF

Informationsbeschaffung und –zugriff

• **Literaturquellen**

Literaturdatenbanken

- Kriterien zur Auswahl der erfaßten Literatur
 - a, erfaßte Zeitschriften b, methodische Kriterien
- Konventionelle Quellen
 - EBM-Zeitschriften / EBM-basierte Lehrbücher
- Zeitaufwand / Aktualität

• **Gemeinsame Merkmale von Literaturdatenbanken**

- Strukturierung in Feldern (Autoren, Titel, Abstract)
- Schlagwort-Systeme
- Recherche-Systeme für verschiedene Literaturdatenbanken
- Funktionen von Recherche-Systemen
- Syntax in Abhängigkeit von Datenbank und Recherche-System
- Sensitive vs. spezifische Suchen
- Verwaltung von Literatur (von Reference manager bis CATs)

• **Thesaurus**

- Einsatz bei der Suche nach Schlagwörtern
- Synonymsuche für Begriffe
- Hierarchische Organisation
- Einschluß untergeordneter Begriffe
- Problematik der verzögerten Anpassung der Thesauren an aktuelle Entwicklungen

• **Internet**

- Vor- und Nachteile (Qualitätsproblematik)
- Vorhandene Links im Internet

TESTFRAGEN EVIDENCE-BASED MEDICINE

SET A

Vor dem Kurs Nachher

NAME: _____

NUMMER: _____

GRUPPE: _____

Alter: _____ Jahre

Geschlecht:

Weiblich

Männlich

Vorkenntnisse (Mehrfach-Auswahl)

- Medizinstudium
- Zusatzstudium (Public Health, Epidemiologie, etc.)
- EBM-Buch gelesen
- Bereits an einer EBM-Einführung/-Session (weniger als ein Tag) teilgenommen
- Bereits an einem anderen EBM-Kurs (ein Tag oder länger) teilgenommen
- Bereits einen EBM-Kurs als Tutor geleitet

Selbsteinschätzung EBM-Kenntnisse

- Keine
- Wenig

Durchschnitt

Fortgeschritten

Experte

Jahre seit Abschluss des Medizinstudiums (3. Staatsexamen) : _____ Jahre

Derzeitiges Haupt-Tätigkeitsfeld (Einfach-Auswahl)

- Direkte Patientenversorgung (Klinik, Praxis)
 - Chirurgisch
 - Konservativ
 - Sonstiges: _____
- Gesundheitsverwaltung (einschließlich MDK)
- Akademisch (Epidemiologie, Sozialmedizin, Public Health etc.)
- Industrie (Pharma etc.)
- Medizinische Publizistik (Redaktion, Verlag etc.)
- Sonstiges: _____

Derzeitige Position

- Student
- AiP
- Assistenzarzt
- Assistenzarzt mit Facharztbezeichnung
- Oberarzt / Chefarzt
- In eigener Praxis
- Sonstiges: _____

1. FRAGE

In der Notaufnahme stellt sich ein Mann mit seit ca. 24 h bestehenden Schmerzen im rechten Unterbauch vor. Bei der körperlichen Untersuchung ergeben sich keine eindeutigen Zeichen einer Appendizitis. Sie wissen aber, daß in dieser Altersgruppe trotzdem ungefähr jeder 10. Patient mit diesen Beschwerden eine Appendizitis ohne typische Zeichen hat.

Sie veranlassen eine Ultraschall-Untersuchung, weil Sie von der letzten internen Qualitätskontrolle wissen, daß der diensthabende Sonographieur bei der Appendizitis-Diagnostik gute Werte erreicht hat (Wahrscheinlichkeitsverhältnis (=Likelihood-Ratio) bei positiven Befunden 1,8 , bei negativen Befunden 0,2).

In diesem Fall legt sich der Sonographieur darauf fest, daß der Patient eine Appendizitis hat. Als Sie den Chirurgen hinzuziehen wollen, fragt dieser telephonisch, wie hoch Sie jetzt die Wahrscheinlichkeit einschätzen, daß der Patient tatsächlich eine Appendizitis hat.

Sie antworten:

- A Ungefähr 2%
- B Ungefähr 7%
- C Ungefähr 15%
- D Ungefähr 30%
- E Eine Aussage ist vor Eintreffen der Laborbefunde nicht möglich
(Ein unbeschriftetes Fagan-Nomogramm liegt bei)

Antwort 1: _____

2. FRAGE:

Der anwesende Famulant ist von Ihrer Antwort beeindruckt und fragt Sie, wie man ein Wahrscheinlichkeitsverhältnis berechnen kann. Sie erklären ihm, daß die Ultraschall-Diagnosen aus einem bestimmten Zeitraum mit der tatsächlich richtigen Diagnose (aus der Histologie oder Verlaufsbeobachtung) verglichen wurde. Da Ihnen die Zahlen nicht mehr erinnerlich sind, demonstrieren Sie dies an einem Zahlenbeispiel:

		Tatsächliche Diagnose Appendizitis	
		Ja	Nein
Sonographische Diagnose Appendizitis	Ja	90	10
	Nein	20	110

In diesem Beispiel ist das Wahrscheinlichkeitsverhältnis für einen positiven Befund:

A,
$$0,09 = \frac{\frac{10}{(10 + 110)}}{1 - \frac{20}{(110 + 90)}}$$

B,
$$0,10 = \frac{\frac{10}{(110 + 10)}}{\frac{90}{(90 + 20)}}$$

C,
$$1,54 = \frac{\frac{20}{(20 + 110)}}{1 - \frac{90}{(90 + 10)}}$$

D,
$$9,82 = \frac{\frac{90}{(90 + 20)}}{\frac{10}{(10 + 110)}}$$

E, Das Wahrscheinlichkeitsverhältnis läßt sich aus diesen Angaben nicht berechnen

Antwort 2: _____

3. FRAGE:

Der Famulant ist jetzt völlig begeistert und behauptet über die Patienten in Ihrem Zahlenbeispiel (Frage 2):

- 1) Ohne jegliche weitere Information könne man feststellen, daß die Beispielpatienten mit pathologischen Ultraschallbefund mit 90-prozentiger Wahrscheinlichkeit tatsächlich eine Appendizitis haben.
- 2) Ohne jegliche weitere Information könne man feststellen, daß für die Beispielpatienten mit normalem Ultraschall-Befund die Wahrscheinlichkeit eines Falschbefundes 18% beträgt ($20/(20+90)$)
- 3) Der positive prädiktive Wert berechnet sich direkt aus dem Quotienten der Wahrscheinlichkeitsverhältnisse ($90\% = 0,09/0,01$)

Sie sagen ihm:

- A Alle Aussagen (1, 2, 3) sind falsch.
- B Die erste Aussage (1) ist richtig, die beiden anderen Aussagen (2, 3) sind falsch.
- C Die zweite und die dritte Aussage (2, 3) sind richtig, die erste Aussage (1) ist falsch.
- D Die erste und die dritte Aussage (1, 3) sind richtig, die zweite Aussage (2) ist falsch.
- E Alle Aussagen (1, 2, 3) sind richtig.

Antwort 3: _____

4. FRAGE:

Ein Pharmavertreter besucht Sie in Ihrer Praxis und stellt Ihnen den neuesten Lipidsenker vor, der in einer großen doppel-blind randomisiert kontrollierten Studie an gesunden Mitarbeitern des Pharmakonzerns eine 50%ige Reduktion des Risikos an einem Herzinfarkt zu versterben erreicht hat :

- 4.000 Personen wurden behandelt, von diesen verstarben 4 (0,1%) an einem Herzinfarkt.
- Von den 4.000 nicht behandelten Kontroll-Personen verstarben 8 (0,2%).

Der Pharmavertreter empfiehlt Ihnen daher, alle Patienten mit dem neuen Medikament zu behandeln. Sie wollen zumindest einem Patienten in dieser Praxis so das Leben retten, stellen aber fest, daß Sie

- A dazu 1000 Patienten behandeln müssen ($= 1 / (0,2\% - 0,1\%)$)
- B dazu 2000 Patienten behandeln müssen ($= 8000 / 4$)
- C dazu 4000 Patienten behandeln müssen ($= 4 * (1 / 0,1\%)$)
- D dazu 8000 Patienten behandeln müssen ($= 4000 * 2$)
- E die Zahl der Patienten, die Sie dazu behandeln müssen, nicht berechnen können

Antwort 4: _____

5. FRAGE:

Sie freuen sich über diese Gelegenheit, der Menschheit zu helfen, und rufen einen Kollegen an, um ihm die frohe Botschaft mitzuteilen. Ihr Kollege hat aber bereits die Veröffentlichung der Studiendaten (Frage 4) gelesen und weist Sie auf die Häufung von Lungenödemen unklarer Genese in der Behandlungsgruppe hin:

- 7 Fälle in der Behandlungsgruppe
- 2 Fälle in der Kontrollgruppe.

Sie rechnen nach und stellen fest, daß ein zusätzliches Lungenödem bei

- A einer Zahl von 1000 behandelten Patienten ($= 2000 * 0,2\%$) zu erwarten ist.
- B einer Zahl von 571 behandelten Patienten ($= 4000 / 7$) zu erwarten ist
- C einer Zahl von 800 behandelten Patienten ($= 1 / (5 / 4000)$) zu erwarten ist
- D einer Zahl von 2000 behandelten Patienten ($= 2 * (1 / 0,1\%)$) zu erwarten ist
- E einer aus diesen Daten nicht berechenbaren Anzahl von behandelten Patienten zu erwarten ist

Antwort 5: _____

6. FRAGE:

Eine Patientin mit chronischen Kopfschmerzen hat kürzlich in ihrer Fernsehzeitschrift gelesen, daß Formaldehyd-belastete Wohnräume chronische Kopfschmerzen verursachen. Sie möchte, daß die Krankenkasse nun ihren Umzug finanziert. Daher fragt sie, ob der Zusammenhang zwischen chronischen Kopfschmerzen und Formaldehyd in der Raumluft wissenschaftlich bewiesen sei. Sie machen eine entsprechende Literaturrecherche und finden verschiedene Studien zu diesem Thema.

Welches Studien-Design halten Sie für die Untersuchung dieser Frage am geeignetsten?

- A Prävalenzstudie
- B ökologische Studie
- C Fall-Kontroll-Studie
- D Prospektive randomisiert-kontrollierte Studie
- E Fall-Serie

Antwort 6: _____

7. FRAGE:

Im einzelnen sind bei den von Ihnen gefundenen Studien die folgenden Vorgehensweisen gewählt worden.

Welche halten Sie für am besten geeignet um zu prüfen, ob tatsächlich ein Zusammenhang zwischen Kopfschmerz und Formaldehyd in der Raumluft besteht?

A Bei 100 Patienten mit Kopfschmerzen aus einer Spezialambulanz und 100 Patienten aus Allgemeinarztpraxen ohne Kopfschmerz, die von Alter, Geschlecht und Einkommen den Kopfschmerzpatienten entsprechen, wird der Formaldehydgehalt in den Wohnräumen gemessen.

Vergleich des mittleren Formaldehydgehaltes in beiden Gruppen.

B Befragung von 500 Patienten einer Ambulanz für Umweltmedizin, ob sie 1.) Kopfschmerzen haben und 2.) meinen, daß ihre Wohnräume formaldehydbelastet sind.

Vergleich der Häufigkeiten der Antwort zu 2.) (Formaldehydbelastete Wohnräume) bei Patienten mit Kopfschmerzen vs. ohne Kopfschmerzen

C Befragung von Mietern einer Wohnungsbaugesellschaft mit einem Kopfschmerzfragebogen. Gleichzeitige Durchführung eines Haut-Tests auf Formaldehydallergie.

Vergleich der Häufigkeit von Kopfschmerzen bei Mietern mit Formaldehydallergie vs. übrige Mieter.

D Zweimalige Messung der Formaldehydkonzentration im Blut von Kopfschmerzpatienten aus einer Schmerzambulanz im Abstand von einem Jahr.

Vergleich der Werte zu Beginn mit denen am Ende der Beobachtung.

E Befragung von Personen, die neu in vermutlich formaldehydbelastete Wohnräume einer städtischen Wohnungsbaugesellschaft ziehen und von langjährigen Mietern solcher Wohnungen mit einem Kopfschmerzfragebogen.

Vergleich der Häufigkeit von Kopfschmerzen zwischen neuen und langjährigen Mietern.

Antwort 7: _____

8. FRAGE:

In einer großen Studie wurde untersucht, ob eine neue Schlankheitspille die kardiovaskuläre Mortalität senkt.

An der Studie nahmen 1000 übergewichtige Patienten teil. Nach dem Zufallsprinzip wurden 500 Patienten in die Behandlungsgruppe und 500 in die Kontrollgruppe eingeteilt. Die Behandlungsgruppe erhielt über ein Jahr die Schlankheitspille, die Kontrollgruppe eine genauso aussehende Milchzuckerpille (Plazebo). Die Patienten der Behandlungsgruppe wurden monatlich auf das Auftreten von Nebenwirkungen untersucht, die Patienten der Kontrollgruppe hingegen nur halbjährlich.

In der Behandlungsgruppe starben 10 Patienten weniger an kardiovaskulären Ereignissen als in der Kontrollgruppe ($p < 0,01$).

Sie überlegen, ob die Qualität der Studie Sie überzeugt. Folgende Feststellung hierzu trifft am ehesten zu:

- A Die Studie wurde nicht prospektiv durchgeführt
- B Die Studie wurde nicht randomisiert durchgeführt
- C Die Studie wurde nicht doppelblind durchgeführt
- D In der Studie wurde nicht der für die Fragestellung wirklich interessierende Endpunkt sondern ein Surrogatendpunkt untersucht.
- E In dieser prospektiven randomisiert-kontrolliert doppelblind durchgeführten Studie wurde der für die Fragestellung interessierende Endpunkt untersucht.

Antwort 8: _____

9. FRAGE:

- 1.) Aus den vorhandenen Angaben (Frage 8) kann man errechnen, wieviele Patienten wie in der Studie mit Schlankheitspille behandelt werden müssen, um einen zusätzlichen kardiovaskulären Todesfall zu verhindern.
- 2.) Für die Berechnung der Zahl der Patienten, die behandelt werden müssen, um einen zusätzlichen Todesfall zu verhindern, muss man die relative Verringerung des Sterbe-Risikos der Behandlungsgruppe gegenüber der Kontrollgruppe kennen.
- 3.) In den vorhandenen Angaben (Frage 8), ist die relative Verringerung des Sterbe-Risikos der Behandlungsgruppe gegenüber der Kontrollgruppe (relative Risikoreduktion) angegeben.

- A Alle Aussagen (1, 2, 3) sind falsch.
- B Die erste Aussage (1) ist richtig, die beiden anderen Aussagen (2, 3) sind falsch.
- C Die zweite und die dritte Aussage (2, 3) sind richtig, die erste Aussage (1) ist falsch.
- D Die erste und die dritte Aussage (1, 3) sind richtig, die zweite Aussage (2) ist falsch.
- E Alle Aussagen (1, 2, 3) sind richtig.

Antwort 9: _____

10. FRAGE:

Sie sind Internistin. Ihre 63jährige Patientin hat als Zufallsbefund eine 70%ige Stenose der linken A. carotis. Sie fragen sich, ob dies eine Indikation zur Überweisung an den Neurochirurgen zur Endarterektomie darstellt und finden eine Studie, die bei asymptomatischen Patienten mit 70%iger Stenose (vergleichbar Ihrer Patientin) bei 5 Jahren Nachbeobachtung keinen Nutzen durch eine Operation gegenüber konservativer Behandlung findet.

In einer Untergruppenanalyse (insgesamt 13 Untergruppen) zeigt sich nach Berücksichtigung des Risikofaktorstatus zu Studienbeginn, daß nur diejenigen Frauen, die das erste Jahr ohne Insult oder TIA überstanden haben, im Verlauf der nachfolgenden 4 Jahre ein statistisch signifikanten Nutzen ($p < 0.03$) von der Endarterektomie hatten.

Welche Aussagen trifft zu:

- A Das signifikante Ergebnis belegt den Operationsnutzen bei Frauen und reicht allein aus, um die Operationsindikation zu rechtfertigen.
- B In der Untergruppen-Analyse wurde für andere Risikofaktoren korrigiert, was meistens zu irreführenden Schlußfolgerungen führt.
- C Mit Untergruppen-Analysen läßt sich die Ausbeute an verlässlichen Ergebnissen aus randomisiert kontrollierten Studien maximieren.
- D Mit zunehmender Zahl von nachträglich gebildeten Untergruppen steigt die Gefahr, daß in einer Untergruppe irrtümlich ein Nutzen gefunden wird, der in Wirklichkeit nicht existiert.
- E Die Untergruppenanalyse beweist in diesem Fall, daß asymptomatische Carotisstenosen für Frauen gefährlicher sind als für Männer.

Antwort 10: _____

11. FRAGE:

In Fortbildung über die Auswirkung von Lipidsenkern nach Myokardinfarkt werden für unterschiedliche Medikamente folgende randomisierte kontrollierte Studien vorgestellt, die alle bei Frauen mittleren Alters ($\bar{\text{O}}$ Alter: 55 J.) mit eingestellter Hypertonie ($\bar{\text{O}}$ RR: 155/98 mmHg) und Adipositas über einen Zeitraum von fünf Jahren gegen Placebo getestet wurden:

- 1) In der bolivianischen Studie senkte Therapie II das Risiko für einen tödlichen Infarkt um 25%.
- 2) In einer argentinischen Studie senkte die Therapie I das Risiko tödlicher Herzinfarkte um 30%.
- 3) In einer chilenischen Studie verstarben in der Gruppe mit Behandlung III 3% der Patienten und in der Kontrollgruppe 4% der Patienten an einem Herzinfarkt.

Welche Aussage trifft zu?

- A Therapie 1 ist zu bevorzugen da diese Therapie das Herzinfarkt-Risiko am stärksten senkt.
- B Therapie 3 ist zu bevorzugen da bei ihr der Anteil der Patienten die von der Behandlung profitieren am größten ist
- C Bei Therapie 2 und 3 ist die relative Reduktion des Risikos eines tödlichen Infarktes gleich groß
- D Bei Therapie 1 und 3 ist die relative Reduktion des Risikos eines tödlichen Infarktes gleich groß
- E Für keine der Therapien wird das Risiko unbehandelter Patienten (Control-Event-Rate) angegeben

Antwort 11: _____

12. FRAGE:

In einer gastro-enterologischen Poliklinik besteht eine Prävalenz von 30% für Kolonkarzinom. 1000 aufeinanderfolgende Patienten werden in eine Studie für einen neuen, nicht invasiven, wenig belastenden diagnostischen Test zum Nachweis von Kolon-Karzinomen aufgenommen. Mit Hilfe des Tests werden 630 Patienten als richtig negativ erkannt, d.h. richtigerweise tumorfrei. Die Anzahl an falsch negativen und falsch positiven Patienten ist identisch.

	Goldstandard positiv	Goldstandard negativ	
Test positiv			
Test negativ			

(4-Felder-Tafel für die eigene Berechnung des Ergebnisses)

Welche 4-Felder-Tafel entspricht diesen Angaben?

4-Felder-Tafel A

	Goldstandard positiv	Goldstandard negativ	
Test positiv	230	70	300
Test negativ	70	630	700
	300	700	1000 Patienten

4-Felder-Tafel B

	Goldstandard positiv	Goldstandard negativ	
Test positiv	300	0	300
Test negativ	0	700	700
	300	700	1000 Patienten

4-Felder-Tafel C

	Goldstandard positiv	Goldstandard negativ	
Test positiv	270	100	370
Test negativ	30	600	630
	300	700	1000 Patienten

4-Felder-Tafel D

	Goldstandard positiv	Goldstandard negativ	
Test positiv	670	30	700
Test negativ	30	270	300
	700	300	1000 Patienten

4-Felder-Tafel E

	Goldstandard positiv	Goldstandard negativ	
Test positiv	300	70	370
Test negativ	70	560	630
	370	630	1000 Patienten

Antwort 12: _____

FRAGE 13:

Vor einiger Zeit hatten Sie eine 40 jährige Patientin wegen eines getasteten Knotens in der Brust zum Gynäkologen überwiesen. Der Knoten wurde nach Punktion als gutartig eingestuft. Die Patientin bringt den Befund mit und möchte von Ihnen wissen, ob die Zyste bedeutet, daß sie besonders gefährdet ist, Brustkrebs zu entwickeln. Sie finden dazu mehrere Studien.

Welche Studie ist am ehesten geeignet die prognostische Bedeutung von gutartigen Zysten bei Frauen aus einer Normalbevölkerung zu beurteilen?

- A Studie aus der gynäkologischen Ambulanz einer Pariser Universitätsklinik:
Von 1996 - 1998 wurden alle Frauen mit Mamma-Karzinom nach Vorbefunden von tastbaren Mammazysten befragt. Zum Vergleich wurden Frauen ohne Mamma-Karzinom auf tastbare Brustzysten untersucht. Frauen mit Mamma-Karzinom hatten 20% weniger Zysten als Frauen ohne Mamma-Karzinom.
- B Studie aus einer Universitätsklinik in Boston:
Die Mammographiebilder von 1500 Frauen mit Mamma-Karzinom aus den letzten 10 Jahren wurden nachträglich auf das Vorhandensein von Zysten untersucht Dabei fanden sich bei 30% der Frauen auch größere Zysten.
- C Studie aus der gynäkologischen Ambulanz des einzigen Überweisungszentrums in Ostschottland für Frauen mit Mamma-Problemen:
Alle Frauen, die wegen tastbarer Zysten abgeklärt worden waren, wurden 10 Jahre später nachuntersucht, bzw. im Krebsregister und Sterberegister gesucht. Verglichen mit der Normalbevölkerung war die Häufigkeit für Mamma-Karzinom doppelt so hoch.
- D Studie aus der Pathologie einer Spezialklinik im Ruhrgebiet:
Alle Histologien von Mamma-Gewebe aus den letzten 10 Jahren wurden erfaßt. Bei 11 % der Patientinnen bei denen eine gutartige Zyste diagnostiziert wurde, wurde innerhalb des Erfassungszeitraums auch ein Mammakarzinom histologisch gesichert.
- E Multizentrische Studie in mehreren Kreiskrankenhäusern:
Erfahrene Mamma-Chirurgen wurden befragt, wie oft nach ihrer Erfahrung bei Mamma-Karzinom-Patientinnen in der Vorgeschichte gutartige Zysten diagnostiziert wurden. Der Mittelwert der Angaben lag bei 37%.

Antwort 13: _____

FRAGE 14:

In den deutschsprachigen Ländern wird der frei flottierende Thrombus bei einer tiefen Beinvenenthrombose wegen der mutmaßlich höheren Emboliegefahr besonders gefürchtet. Sie möchten wissen, ob Patienten mit einem frei flottierenden Thrombus ein höheres Risiko für eine Lungenembolie haben als Patienten mit wandständigem Thrombus.

- 1) Dies ist eine Frage über Prognose
- 2) Dies ist eine Frage über Nebenwirkungen
- 3) Diese Frage wird am besten in einer Fall-Kontrollstudie untersucht
- 4) Diese Frage wird am besten in einer Kohortenstudie untersucht.
- 5) Diese Frage wird am besten in einer randomisierten kontrollierten Studie untersucht

Welche Aussage trifft zu

- A 1 und 3 sind richtig
- B 1 und 4 sind richtig
- C 1 und 5 sind richtig
- D 2 und 4 sind richtig
- E 2 und 5 sind richtig

Antwort 14: _____

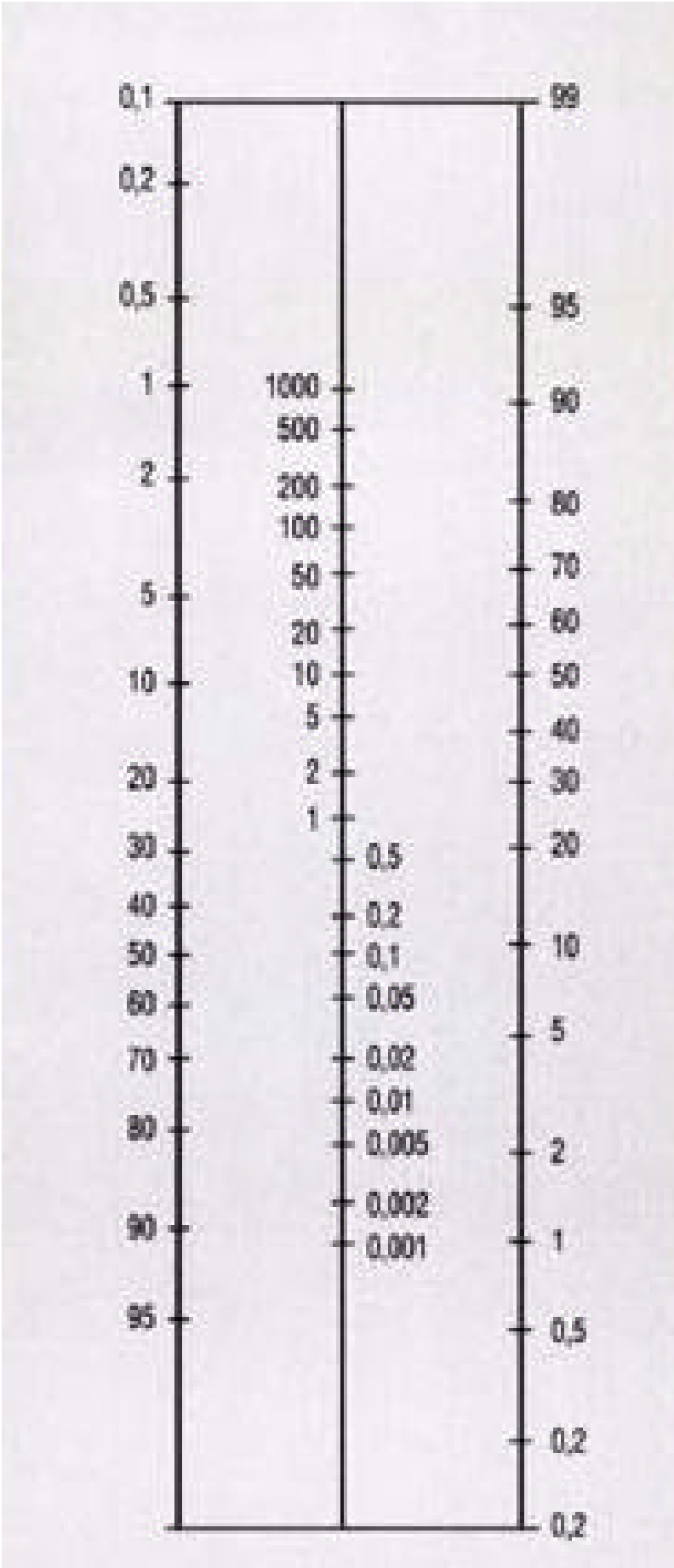
FRAGE 15:

Die Meta-Analyse ist ein Studientyp, der in letzter Zeit zunehmende Bedeutung erlangt hat.

Welche Aussage trifft zu?

- A Durch die Technik der Meta-Analyse ist die Durchführung großer Studien mit vielen Patienten (Mega-Trials) weniger bedeutsam geworden.
- B In systematischen Untersuchungen konnte nachgewiesen werden, daß Studien, die in der jeweiligen Landessprache veröffentlicht wurden, in der Qualität schlechter sind als die Studien in englischen Zeitschriften. Aus diesem Grunde genügt für eine Meta-Analyse die Berücksichtigung der englischsprachigen Literatur.
- C Studien mit größeren Patientenzahlen finden häufig auch einen größeren Therapie-Effekt als Studien mit einer kleineren Patientenzahl.
- D In einer Meta-Analyse randomisierter Studien die insgesamt eine hohe Patientenzahl umfaßt, kann die Wirksamkeit von Maßnahmen genauer geschätzt werden als in kleineren Einzelstudien .
- E Bei der Meta-Analyse können Unterschiede in der Fragestellung oder den Patientenpopulationen der einzelnen Studien mit Hilfe der Statistik ausgeglichen werden.

Antwort 15: _____



TESTFRAGEN EVIDENCE-BASED MEDICINE

SET B

Vor dem Kurs Nachher

NAME: _____

NUMMER: _____

GRUPPE: _____

Alter: _____ Jahre

Geschlecht:

Weiblich

Männlich

Vorkenntnisse (Mehrfach-Auswahl)

Medizinstudium

Zusatzstudium (Public Health, Epidemiologie, etc.)

EBM-Buch gelesen

Bereits an einer EBM-Einführung/-Session (weniger als ein Tag) teilgenommen

Bereits an einem anderen EBM-Kurs (ein Tag oder länger) teilgenommen

Bereits einen EBM-Kurs als Tutor geleitet

Selbsteinschätzung EBM-Kenntnisse

Keine

Wenig

Durchschnitt

Fortgeschritten

Experte

Jahre seit Abschluss des Medizinstudiums (3. Staatsexamen) : _____ Jahre

Derzeitiges Haupt-Tätigkeitsfeld (Einfach-Auswahl)

Direkte Patientenversorgung (Klinik, Praxis)

Chirurgisch

Konservativ

Sonstiges: _____

Gesundheitsverwaltung (einschließlich MDK)

Akademisch (Epidemiologie, Sozialmedizin, Public Health etc.)

Industrie (Pharma etc.)

Medizinische Publizistik (Redaktion, Verlag etc.)

Sonstiges: _____

Derzeitige Position

Student

AiP

Assistenzarzt

Assistenzarzt mit Facharztbezeichnung

Oberarzt / Chefarzt

In eigener Praxis

Sonstiges: _____

FRAGE 1

Sie sind Stationsarzt auf der gynäkologischen Station eines Krankenhauses auf dem Land (Grundversorgung). Einer 46-jährigen Patientin wurde in dieser Woche ein kleines Mammakarzinom (pT1 pN0 M0) operativ entfernt. Die Patientin ist aufgrund der erst kürzlich mitgeteilten Diagnose völlig aufgelöst und möchte nun von Ihnen wissen, wie lange sie ohne weitere Therapiemaßnahmen noch zu leben hat. Sie sind sich nicht sicher, ob und in welchem Maße die Lebenserwartung der Patientin nach erfolgreicher OP ohne Nachbehandlung eingeschränkt ist. Sie wollen die Frage anhand geeigneter Studien beantworten.

Welcher Studientyp eignet sich am besten zur Beantwortung dieser Frage:

- A Fall-Kontroll-Studie
- B Querschnittsstudie
- C Kohortenstudie
- D Fall-Serie
- E Prävalenzstudie

Antwort 1: _____

2. FRAGE:

Bei Ihrer Literaturrecherche stoßen Sie auf folgende Studien. Welche ist am besten zur Beantwortung der Frage der Patientin (s. Frage 1) geeignet?

- A In einer großen Multi-Center-Studie wurden 3600 Frauen nach operativer Entfernung eines Mammakarzinoms in zwei Gruppen randomisiert. Eine Gruppe wurde nachbestrahlt, die andere erhält eine adjuvante Chemotherapie. Die Überlebensraten beider Gruppen werden angegeben.
- B In einer süddeutschen Kleinstadt wurden alle Frauen über 40 befragt, ob sie schon eine Brustkrebs-OP hinter sich haben. Für die Gruppe der Frauen mit Brustkrebs-OP in der Anamnese wird aus der Altersverteilung die Sterberate für die verschiedenen Tumorstadien kalkuliert.
- C Der mittlerweile pensionierte Leiter Ihrer Abteilung hat über Jahre Fälle von Frauen mit T1-Karzinomen, die er operiert hat, gesammelt und seine Erfahrungen wurden zu seinem 65-ten in einem Sonderband publiziert. Keine einzige Patientin ist vorzeitig gestorben.
- D In einem deutschen Stadtstaat haben sich alle gynäkologischen Kliniken zu einer Studiengruppe zusammengeschlossen. Alle Frauen, bei denen operativ eine bösartige Geschwulst der Brust entfernt wurde, werden mit ihrem Tumorstadium zum Zeitpunkt der OP registriert. In jährlichen Abständen wird überprüft, ob die Frauen noch am Leben sind. Die Überlebensrate wird nach Tumorstadium gruppiert angegeben.
- E Das pathologische Institut einer renommierten deutschen Universitätsklinik hat alle Sektionsfälle, bei denen in der Vorgeschichte ein Mammakarzinom operiert wurde registriert. Aus der Differenz zwischen OP-Datum und Todesdatum wird die Überlebensrate berechnet und gruppiert nach Tumorstadien angegeben.

Antwort 2: _____

3. FRAGE:

Die Patientin hat in einer (amerikanischen) Frauen-Zeitschrift eine Anzeige für Tamoxifen gelesen. Sie möchte von Ihnen wissen, ob es sich bei Tamoxifen um ein Medikament mit einer nachgewiesenen Wirkung handelt.

- 1.) Für den Nachweis einer Medikamentenwirkung ist eine doppel-blinde randomisierte kontrollierte Studie ein geeigneter Studien-Typ.
- 2.) Die Randomisierung sollte zu einer ungefähr gleichen Verteilung der Patienteneigenschaften zwischen Behandlungs- und Kontrollgruppe führen.
- 3.) Durch die Randomisierung sollten auch unbekannte Faktoren, die für die Heilung bedeutsam sein können, in beiden Gruppen mit gleicher Wahrscheinlichkeit wirksam werden.

- A Alle Aussagen (1, 2, 3) sind falsch.
- B Die erste Aussage (1) ist richtig, die beiden anderen Aussagen (2, 3) sind falsch.
- C Die zweite und die dritte Aussage (2, 3) sind richtig, die erste Aussage (1) ist falsch.
- D Die erste und die dritte Aussage (1, 3) sind richtig, die zweite Aussage (2) ist falsch.
- E Alle Aussagen (1, 2, 3) sind richtig.

Antwort 3: _____

4. FRAGE:

In einer Studie zur Wirksamkeit von Tamoxifen bei Brustkrebs wurden in 37 amerikanischen Kliniken insgesamt 10.000 Frauen direkt nach operativer Entfernung eines Mammakarzinoms eingeschlossen. In jede Klinik wurde eine Liste von mit einem Zufallsgenerator erzeugten Zahlen ausgehängt. Bei Einschluß einer Patientin wurde die nächste noch nicht benutzte Zahl der Liste dieser Patientin zugeordnet und dann durchgestrichen. Handelte es sich um eine gerade Zahl wurde die Patientin für drei Jahre mit 2x30 mg Tamoxifen täglich behandelt. Bei einer ungeraden Zahl wurde die entsprechende Patientin über den gleichen Zeitraum beobachtet aber nicht behandelt. Am Ende des Beobachtungszeitraums von drei Jahren wurde festgestellt, wie viele Patientinnen in der behandelten und in der nicht behandelten Gruppe jeweils verstorben waren. In der behandelten Gruppe verstarben innerhalb von drei Jahren 800 von 5000 Patientinnen, in der nicht behandelten Gruppe 1000 von 5000.

Die Studie hatte folgende Eigenschaften

- 1) randomisiert
- 2) verdeckt randomisiert (concealed allocation)
- 3) doppel-blind
- 4) kontrolliert
- 5) multi-zentrisch

Welche Eigenschaften hat die Studie?

- A 1, 2 und 5 sind richtig
- B 1, 3 und 5 sind richtig
- C 1, 4 und 5 sind richtig
- D 2 und 3 sind richtig
- E 4 und 5 sind richtig

Antwort 4: _____

5. FRAGE:

Welche Aussage(n) zur Beeinflussung des Sterblichkeitsrisikos durch die Behandlung trifft in dieser Studie zu

- 1) In der behandelten Gruppe sank das Risiko zu versterben um 20% des Risikos der nicht behandelten Gruppe (relative Risikoreduktion =20% = $(1000-800)/1000$)
- 2) In der behandelten Gruppe sank das Risiko zu versterben um 80% des Risikos der nicht behandelten Gruppe (relative Risikoreduktion =80% = $(800/1000)$)
- 3) In der behandelten Gruppe lag das Sterblichkeitsrisiko um 2 Prozentpunkte unter dem der nicht behandelten Gruppe (absolute Risikoreduktion =2% = $(1000-800)/10.000$)
- 4) In der behandelten Gruppe lag das Sterblichkeitsrisiko um 4 Prozentpunkte unter dem der nicht behandelten Gruppe (absolute Risikoreduktion =4% = $(1000-800)/5.000$)
- 5) In der behandelten Gruppe lag das Sterblichkeitsrisiko um 10 Prozentpunkte unter dem der nicht behandelten Gruppe (absolute Risikoreduktion =10% = $1000/10.000$)

- A 1 und 3 sind richtig
- B 1 und 4 sind richtig
- C 1 und 5 sind richtig
- D 2 und 4 sind richtig
- E 2 und 5 sind richtig

Antwort 5: _____

6. FRAGE:

Angenommen in dieser Studie (Frage 5) habe die relative Risikoreduktion 50% und die absolute Risikoreduktion 2% betragen.

Wie viele Patientinnen müßten dann mit Tamoxifen behandelt werden, um einen zusätzlichen Todesfall zu verhindern?

- A $48 = 50 - 2$
- B $50 = 1 / 2\%$
- C $52 = 50 + 2$
- D $100 = 5000 \times 2\%$
- E Für diese Berechnung sind weitere Angaben notwendig

Antwort 6: _____

7. FRAGE:

In einer großen multi-zentrischen doppel-blinden Studie wurde untersucht, ob durch Dauermedikation mit einem Statin die Gefahr von Schlaganfällen gesenkt werden kann. Behandlungs- und Placebogruppe waren genau gleich groß.

In der Placebo-Gruppe erlitten 25% der Patienten einen Schlaganfall.

Die absolute Risiko-Reduktion betrug 2,4% (95% Konfidenzintervall = 0,5% - 20%)

Bitte prüfen Sie folgende Feststellungen

- 1) Durch die Einnahme des Statins (Behandlung) wurde das Risiko einen Schlaganfall zu erleiden um etwa 10% reduziert.
- 2) Durch die Einnahme des Statins (Behandlung) wurde bei etwa 2,4% der Patienten ein Schlaganfall verhindert
- 3) Der Unterschied zwischen Placebo- und Statin-(Behandlungs-)Gruppe ist statistisch signifikant

Welche Aussage(n) treffen zu

- A keine Aussage ist richtig
- B 1 ist richtig
- C 2 ist richtig
- D 2 und 3 sind richtig
- E Alle Aussagen sind richtig

Antwort 7: _____

8. FRAGE:

In der Notaufnahme Ihres Krankenhauses soll ein neuer Test für den Nachweis von Thrombosen eingeführt werden. Der D-Dimer Bedside-Test hat nur zwei mögliche Ergebnisse: Positiv (=Thrombose) oder Negativ (=Keine Thrombose). Die Sensitivität beträgt 99%, die Spezifität liegt bei 60%.

- 1.) Ein diagnostischer Test mit einer sehr hohen Sensitivität ist gut für den Ausschluß der untersuchten Erkrankung (hier: Thrombose) geeignet
- 2.) Eine hohe Sensitivität bedeutet, daß der diagnostische Test die gesuchte Erkrankung mit hoher Wahrscheinlichkeit findet wenn sie tatsächlich vorhanden ist.
- 3.) Bei einem diagnostischen Tests mit einer sehr hohen Sensitivität ist die Wahrscheinlichkeit sehr niedrig, daß trotz eines negativen Test-Ergebnisses (hier: Keine Thrombose) die gesuchte Erkrankung (hier: Thrombose) trotzdem vorliegt

- A Alle Aussagen (1, 2, 3) sind falsch.
- B Die erste Aussage (1) ist richtig, die beiden anderen Aussagen (2, 3) sind falsch.
- C Die zweite und die dritte Aussage (2, 3) sind richtig, die erste Aussage (1) ist falsch.
- D Die erste und die dritte Aussage (1, 3) sind richtig, die zweite Aussage (2) ist falsch.
- E Alle Aussagen (1, 2, 3) sind richtig.

Antwort 8: _____

9. FRAGE:

Die Eigenschaften des D-Dimer Bedside-Tests wurde in einer Studie untersucht. Es wurden bei einer Prävalenz von 30% für Thrombosen 1000 aufeinanderfolgende Patienten in die Studie aufgenommen. Mit Hilfe des Tests werden 630 Patienten als richtig negativ erkannt, d.h. richtigerweise thrombosefrei. Die Anzahl an falsch negativen und falsch positiven Patienten ist identisch.

	Goldstandard positiv	Goldstandard negativ	
Test positiv			
Test negativ			

(4-Felder-Tafel für die eigene Berechnung des Ergebnisses)

Welche 4-Felder-Tafel entspricht diesen Angaben?

4-Felder-Tafel A

	Goldstandard positiv	Goldstandard negativ	
Test positiv	230	70	300
Test negativ	70	630	700
	300	700	1000 Patienten

4-Felder-Tafel B

	Goldstandard positiv	Goldstandard negativ	
Test positiv	300	0	300
Test negativ	0	700	700
	300	700	1000 Patienten

4-Felder-Tafel C

	Goldstandard positiv	Goldstandard negativ	
Test positiv	270	100	370
Test negativ	30	600	630
	300	700	1000 Patienten

4-Felder-Tafel D

	Goldstandard positiv	Goldstandard negativ	
Test positiv	670	30	700
Test negativ	30	270	300
	700	300	1000 Patienten

4-Felder-Tafel E

	Goldstandard positiv	Goldstandard negativ	
Test positiv	300	70	370
Test negativ	70	560	630
	370	630	1000 Patienten

Antwort 9: _____

10. FRAGE:

In einer anderen Studie wurde die Eignung des Tests zu Diagnose bzw. Ausschluß einer Lungenembolie untersucht. Bei Untersuchung von 800 Patienten mit und ohne Lungenembolie ergaben sich folgende Ergebnisse:

		Tatsächliche Diagnose Lungenembolie	
		Ja	Nein
Bedside-Test Ergebnis	Lungenembolie	90	210
	Keine Lungenembolie	110	190

Bei diesem Test beträgt die Wahrscheinlichkeit, daß ein Patient, über den Sie sonst nichts wissen, tatsächlich eine Lungenembolie hat, wenn der Test positiv („Lungenembolie“) ausfällt (Positiver prädiktiver Wert):

- A $50\% = (90+110)/(210+190)$
- B $30\% = 90/(90+210)$
- C $45\% = 90/(90+110)$
- D $100\% = (110+190)/(90+210)$
- E $27,5\% = 110/(210+190)$

Antwort 10: _____

11. FRAGE:

Sie benutzen den Test nun zum ersten Mal selbst in der Notaufnahme bei einem älteren Mann, der zwei Tage nach einer Busreise eine rote schmerzhafte Beinschwellung und Fieber entwickelt hat. Sie sind sich überhaupt nicht im klaren, ob er eine Thrombose hat oder nicht. Sie schätzen daher aufgrund Ihres klinischen Eindrucks die Wahrscheinlichkeit, daß eine Thrombose vorliegt, auf 50%.

Der daraufhin durchgeführte D-Dimer Beside-Test ist positiv. Sie wissen aus der oben angeführten Studie, daß das Wahrscheinlichkeitsverhältnis für einen positiven Befund 7,7 und für einen negativen Befund 0,3 beträgt. Daher schätzen Sie nun die Wahrscheinlichkeit, daß eine Thrombose vorliegt, auf

- A Ungefähr 90%
- B Ungefähr 75%
- C Ungefähr 60%
- D Ungefähr 45%
- E Eine Aussage ist vor Durchführung einer Phlebographie nicht möglich
(Ein unbeschriftetes Fagan-Nomogramm liegt bei)

Antwort 11: _____

12. FRAGE:

Der Laborleiter bietet an, daß die D-Dimer Bestimmung auch mit einem ELISA in seinem Labor erfolgen kann. Denn im Labor kann die Konzentration der D-Dimere genau bestimmt und als Zahlenwert angegeben werden. Dadurch soll die Aussagekraft größer sein, als bei dem Bedside-Test, der nur Positiv oder Negativ als Ergebnis liefert. Der Laborleiter unterstreicht seine Argumentation mit einer Studie, in der die D-Dimer-Konzentrationen in fünf Stufen eingeteilt wurden.

Für insgesamt 2579 Patienten mit und ohne Thrombosen ergaben sich folgende Ergebnisse:

Ergebnisstufe	D-Dimer-Konz.	Pat. mit Thrombose		Patienten ohne Thrombose	
Hoch positiv	> 55 mg/dl	474	(59%)	20	(1,1%)
Mäßig positiv	40 – 55 mg/dl	175	(22%)	79	(4,5%)
Neutral	25 – 39 mg/dl	82	(10%)	171	(10%)
Mäßig negativ	10 – 24 mg/dl	30	(3,7%)	168	(9,5%)
Stark negativ	< 10 mg/dl	48	(5,9%)	1332	(75%)
Gesamt:		809	(100%)	1770	(100%)

Das Wahrscheinlichkeitsverhältnis (=Likelihood-Ratio) für ein mäßig positives Ergebnis (40 – 55 mg/dl) ist in dieser Studie:

$$A \quad 4,8 = \frac{\frac{175}{809}}{\frac{79}{1770}}$$

$$B \quad 1,0 = \frac{\frac{175}{1770}}{\frac{79}{809}}$$

$$C \quad 0,02 = \frac{\frac{20}{474}}{\frac{175}{79}}$$

$$D \quad 0,01 = \frac{\frac{30+48}{168+1332}}{\frac{474+175}{20+79}}$$

E Das Wahrscheinlichkeitsverhältnis läßt sich mit diesen Angaben nicht berechnen

Antwort 12: _____

13. FRAGE:

Die Meta-Analyse ist eine wichtige Methode zur Zusammenfassung verschiedener Studien. Welche Eigenschaften verringern in den meisten Fällen die Glaubwürdigkeit einer Meta-Analyse?

- 1) Getrennte Auswertung von Studien mit signifikantem und nicht-signifikantem Ergebnis
- 2) Getrennte Auswertung von randomisierten und nicht-randomisierten Studien
- 3) Beschränkung auf randomisierte Studien
- 4) Beschränkung auf englisch-sprachige Publikationen
- 5) Beschränkung auf Publikationen zu nur einem von verschiedenen denkbaren therapeutischen Vorgehen (z. B. nur Studien zum Einsatz von Tamoxifen bei Brustkrebs, aber nicht Studien zu Gemzytabine bei Brustkrebs)

Die Glaubwürdigkeit einer Meta-Analyse wird durch folgende Eigenschaften meist verringert:

- A 2
- B 1 und 4
- C 2 und 4
- D 1, 2, 4 und 5
- E alle

Antwort 13: _____

14. FRAGE:

Die Benutzung von Funktelefonen wird immer wieder als Gesundheitsrisiko diskutiert. In der HNO-Abteilung Ihres Hauses wurden in kurzer Zeit mehrere Fälle des extrem seltenen (weltweit jährliche Inzidenz unter 1/100.000) extrem bösartigen Ohrmuschelkrebses (kurz: EBOK) bei Handy-Benutzern diagnostiziert. Der Chefarzt fragt Sie daraufhin, ob wissenschaftlich belegt sei, daß Handy-Benutzung EBOK verursacht. Sie starten umgehend eine Literatur-Recherche um diese Frage zu beantworten.

Welches Studien-Design halten Sie für am besten zur Untersuchung dieser Frage geeignet?

- A Prävalenzstudie
- B ökologische Studie
- C Fall-Kontroll-Studie
- D Prospektive randomisiert-kontrollierte Studie
- E Fall-Serie

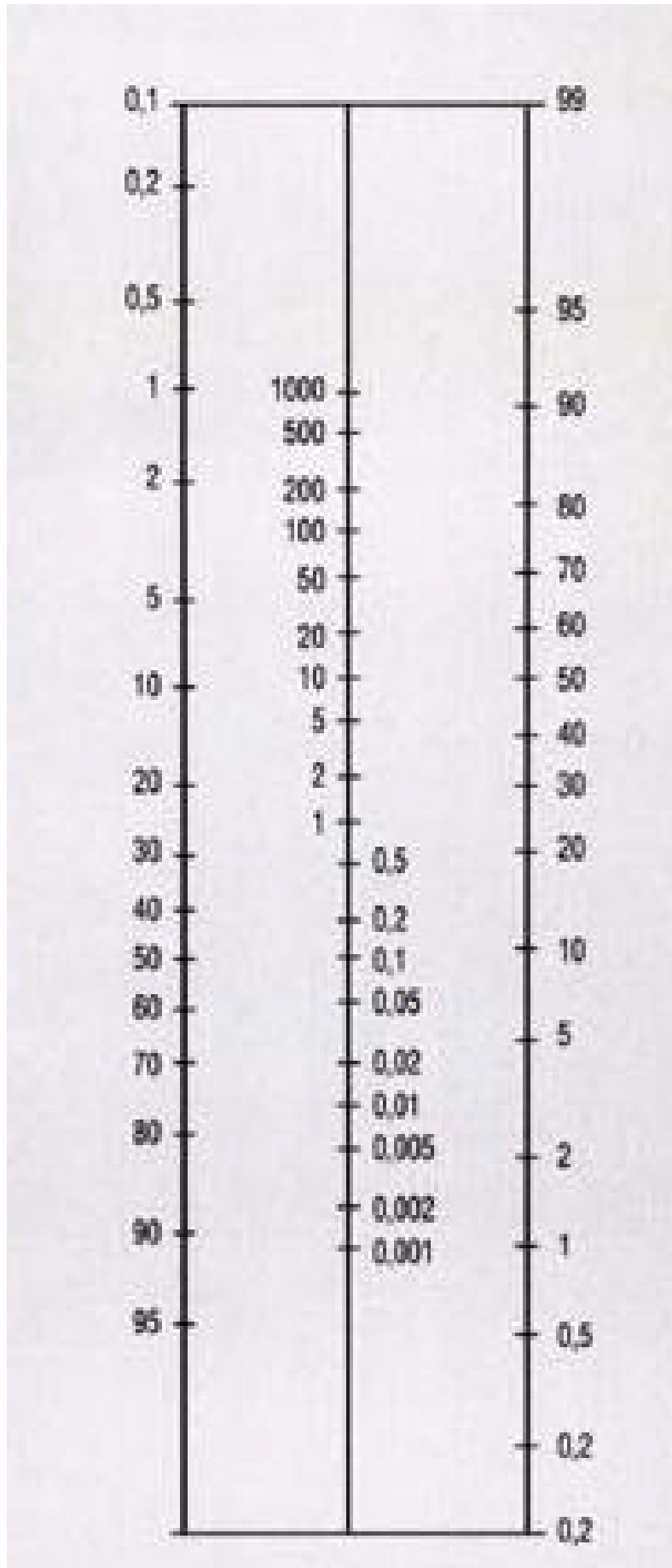
Antwort 14: _____

15. FRAGE:

Sie finden bei Ihrer Literatur-Recherche folgende Studien. Welche ist am am besten zur Beantwortung der Frage geeignet?

- A In Frankreich wurden Fälle von EBOK per Zeitungsannonce und Kontaktierung von Fachärzten gesucht. Jedem EBOK-Patienten wurden vier Patienten aus der Praxis des behandelnden Arztes zugeordnet, die in Geschlecht, Alter und Einkommen mit dem EBOK-Patienten übereinstimmen, aber nicht an EBOK erkrankt sind. Der Anteil an Handy-Benutzern (aktuell und früher) bei EBOK-Patienten und Nicht-EBOK-Patienten wurde verglichen.
- B In einer burmesischen Provinzhauptstadt wurde die Verwaltung (240 Personen) mit Handys ausgerüstet. Handys sind ansonsten in Myanmar (=Burma) strengstens verboten. Nach 5 Jahren wurden alle Verwaltungsangestellten auf EBOK untersucht. Die Häufigkeit von EBOK bei diesen Personen wurde mit der Häufigkeit von EBOK bei normalen Einwohnern der Provinz verglichen.
- C In einem Krebsregister in Ostdeutschland werden alle Krebs-Todesfälle erfaßt. Die Veränderung der Zahl von EBOK-Fällen im Zeitraum 1980 bis 1989 zum Zeitraum 1990 bis 1999 wurden miteinander verglichen.
- D In Italien wurde im Jahr 1996 die Inzidenz von EBOK-Fällen in Nord- und Süditalien erhoben. Der Unterschied in der Inzidenz zwischen beiden Regionen wurde mit dem Unterschied in den Pro-Kopf-Handy-Gesprächsminuten zwischen beiden Regionen verglichen.
- E In einer Befragung der Mehrzahl der Chefarzte deutscher HNO-Kliniken wurde erhoben, zu welchem Anteil nach Meinung der Befragten bei ihren Patienten EBOK auf die Benutzung eines Handys zurückzuführen war.

Antwort 15: _____



Dokumentation des Interviews Interventionsgruppe - 1 Patient

Behandelnder Arzt: _____ / Arzt-ID: _____ Praxis-ID: _____

Tel _____

Fax : _____

Fall

Kontrolle

Patient: _____

Brief-Nr _____

Datum/Brief _____

TBS: _____

Vorgespräch am _____

Besonderes:

Interviewtermin: _____

Interviewer

① Nicola

② Susanne

Ich beziehe mich auf den Arztbrief von Frau / Herrnvom [Datum].....
Wir haben damals die Therapie mit begonnen.

1. Hatten Sie zur Weiterbehandlung des Patienten den Arztbrief inklusive Kommentar gelesen?

ja ⑤	nein ①	keine Angabe ⑨
---------	-----------	-------------------

2. War der Inhalt dieses Kommentars eine Neuigkeit für Sie?

In diesem Fall also, daß(entsprechende Essenz einfügen) ?

ja ⑤	Teilweise ③	Nein ①	keine Angabe ⑨
---------	----------------	-----------	-------------------

3. Um die Therapie-Empfehlungen nachzuvollziehen, fanden Sie den Kommentar

Sehr hilfreich ⑤	Weitgehend hilfreich ④	hilfreich ③	weniger hilfreich ②	Gar-nicht hilfreich ①	keine Angabe ⑨
---------------------	---------------------------	----------------	------------------------	--------------------------	-------------------

4. Wurde die Entlassungsmedikation geändert?

(Ein Produktwechsel innerhalb der Substanzgruppe (z.B. von Ranitidin auf Cimetidin) oder Dosisänderung gilt als Fortsetzung.)
CAVE: a) Empfehlung Medikament zu geben; b) Empfehlung Medikament nicht zu geben.

ja (Es folgen Fragen 5 bis 13) ⑤	Nein (Es folgen Fragen 5, 6, 12, 13) ①	Keine Angabe ⑨
-------------------------------------	--	-------------------

5. Der Patient erhält jetzt bezüglich des oben genannten Problemes.

6. Wie hat der Kommentar diese Entscheidung beeinflusst:

Sehr stark ⑤	Stark ④	Mittelstark ③	Weniger stark ②	Gar nicht ①	Keine Angabe ⑨
-----------------	------------	------------------	--------------------	----------------	-------------------

Qualitätskontrolle durch Studienleitung:

4a, Wurde die Entlassungsmedikation lt. Studienleitung geändert?

(Ein Produktwechsel innerhalb der Substanzgruppe (z.B. von Ranitidin auf Cimetidin) oder Dosisänderung gilt als Fortsetzung.)
CAVE: a) Empfehlung Medikament zu geben; b) Empfehlung Medikament nicht zu geben.

Ja ⑤	Nein ①	Keine Angabe ⑨
---------	-----------	-------------------

Wenn Frage 4 mit Ja beantwortet:

Weshalb wurde die Empfehlung von Ihnen nicht umgesetzt?

Wählen Sie bitte jeweils zwischen ja oder nein.

Bei Mehrfachnennungen markieren Sie bitte den aus Ihrer Sicht wichtigsten Grund!

(Arzt soll sich auf einen wichtigsten Grund festlegen. Falls offensichtlich, daß er sich nicht entscheiden kann, können intern, d.h. ohne daß dem Arzt diese Möglichkeit angeboten wird (!), 2 Felder angekreuzt werden)

	Ja	nein	keine Angabe	Wichtigster Faktor (1 Antwort)
7. Sie haben eine andere Vorstellung über die geeignete Therapie	⑤	①	⑨	⑦
8. Aus medizinischen Gründe, d.h. der Gesundheits-Zustand des Patienten hat sich geändert, Nebenwirkungen sind aufgetreten	⑤	①	⑨	⑧
9. aus nicht-medizinischen Gründen, die den Patienten betrafen (z.B. Wunsch des Patienten; schlechte Compliance, die Zuzahlung war für den Patienten zu hoch)	⑤	①	⑨	⑨
10. aus Budgetgründen	⑤	①	⑨	① ⑩
11. Sonstige Gründe, wenn ja, was::	⑤	①	⑨	① ①

12. Ist für Sie die Empfehlung mit Kommentar ein

sehr starker ⑤	starker ④	mittelstarker ③	Weniger starker ②	kein ①	keine Angabe ⑨
-------------------	--------------	--------------------	----------------------	-----------	-------------------

Eingriff in Ihre Therapie-Freiheit?

13. Wünschen Sie eine Fortsetzung der Kommentare?

ja ⑤	egal ③	nein ①	keine Angabe ⑨
---------	-----------	-----------	-------------------

Soweit meine Fragen zu Pat.; Vielen Dank!

Soziodemographische Faktoren

Zum Schluß möchte ich Ihnen noch einige Fragen zu Ihrer Person stellen.

14. Sie sind

- weiblich // männlich (im Vorfeld ausfüllen, ohne die Frage zu stellen)

15. Wie alt sind Sie?

- Jahre
 keine Angabe

16. In welcher Fachrichtung sind Sie tätig: Als

- Internist mit Hausarztfunktion
 Internist mit Spezialistenfunktion (z.B. Freitext.....)
 Allgemeinarzt
 Andere Fachrichtung
 Keine Angabe

17. Seit wann sind Sie niedergelassen?

Seit

18. Arbeiten Sie in Ihrer Praxis

- alleine
 mit einem oder mehreren Kollegen und Weiterbildungsassistenten

19. Mit wievielen Kollegen arbeiten Sie in Ihrer Praxis zusammen?

Anzahl der Kollegen

- Keine Angabe

Ich habe meine Fragen jetzt beendet. Haben Sie noch Fragen an mich?

Ggf. Freitext -Notizen

Ich danke Ihnen für Ihre Teilnahme an diesem Interview und wünsche Ihnen noch einen schönen Tag.

Institut für Sozialmedizin, Epidemiologie und Gesundheitsökonomie

Direktor: Prof. Dr. Stefan N. Willich, MPH, MBA

Sehr geehrte(r) Frau/Herr Dr.,

wir freuen uns sehr, dass Sie unser Angebot einer wissenschaftlichen Recherche wahrgenommen haben.

Um herauszufinden, ob dieses Angebot Ihrem Informationsbedarf entspricht und ob es einen Einfluss auf Ihre klinische Entscheidungsfindung hatte, möchten wir Sie bitten uns folgende Fragen zu beantworten:

(1) Wie fanden Sie das EbM- Recherche- Angebot insgesamt?

Bitte ankreuzen

sehr schlecht

sehr gut

1 2 3 4 5

(2) Hat das Rechercheergebnis Ihrem Informationsbedarf entsprochen?

Bitte ankreuzen

1. Ja
2. Nein
3. Wenn nein, was fehlte Ihnen?

.....
Bitte eintragen

(3) Wie lange sind Sie schon ärztlich/medizinisch berufstätig?

.....
Bitte eintragen (1Jahr, 2 Jahre etc.)

Anhang Kapitel 5.2.doc

(4) In welcher beruflichen Position stehen Sie zur Zeit?

(Mehrfachantworten möglich)

Bitte ankreuzen

- 1. Ärztin/Arzt im Praktikum
- 2. Ärztin/Arzt in Weiterbildung
- 3. Fachärztin/-arzt
- 4. Ärztin/Arzt in leitender Funktion (Chef-/Oberarzt/-ärztin etc...)
- 5. Als Ärztin/Arzt in Wissenschaft und Forschung
- 6. Sonstiges.....

Bitte eintragen

(5) Wie alt sind Sie?

.....

Bitte eintragen (Jahre)

(6) Bitte kreuzen Sie Ihr Geschlecht an

Bitte ankreuzen

- 1. Weiblich
- 2. Männlich

(7) Haben Sie das Ergebnis für die fachliche Nutzung schnell genug erhalten?

Bitte ankreuzen

- 1. Ja
- 2. Teilweise
- 3. Nein

(8) Ist die Art und Weise, wie Ihre Anfrage beantwortet wurde, für Sie nachvollziehbar und transparent?

Bitte ankreuzen

- überhaupt nicht* *sehr*
- 1 2 3 4 5

(9) Konnten Sie die Empfehlungen der Antwort in Ihren klinischen Alltag integrieren bzw. auf Ihre Patient/-innen anwenden?

Bitte ankreuzen

<i>überhaupt nicht</i>					<i>sehr</i>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5	

(10) Unterschied sich das Ergebnis der Recherche und die daraus abgeleitete Therapieempfehlung von Ihrer eigenen Therapieentscheidung vor Ihrer Anfrage?

Bitte ankreuzen

<i>überhaupt nicht</i>					<i>sehr</i>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5	

(11) Wie schätzen Sie Ihren zeitlichen Aufwand zur Erstellung einer Frage ein?

Bitte ankreuzen

<i>sehr gering</i>					<i>sehr hoch</i>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5	

(12) Haben Sie schon einmal an einem Workshop über Evidence Based Medicine teilgenommen?

Bitte ankreuzen

- | | |
|---------|--------------------------|
| 1. Ja | <input type="checkbox"/> |
| 2. Nein | <input type="checkbox"/> |

(13) Woher beziehen Sie hauptsächlich neue medizinische Informationen?

(Mehrfachantworten möglich)

Bitte ankreuzen

1. Vorgesetzte (Chefarzt/-ärztin, Oberarzt/-ärztin etc)
2. Andere Kollegen/-innen
3. Deutsches Ärzteblatt bzw. andere Kammerblätter (wie z.B. Berliner Ärzte)
4. Fachzeitschriften
5. Kongresse
6. Fort- und Weiterbildungen
7. Pharmazeutische Veranstaltungen,- Beratungen
8. Sonstiges

Bitte eintragen

(14) Haben Sie bisher das Internet für Ihre Suche nach medizinischen Informationen und Literatur genutzt?

Bitte ankreuzen

1. Ja, häufig
2. Ja, gelegentlich
3. Nein, bisher nicht

(15) Wenn ja, wie führen Sie die Literaturrecherche durch?

(Mehrfachantworten möglich)

Bitte ankreuzen

1. Internet-Suchmaschinen
2. Medline (z. B. PubMed, Internet GratefulMed)
3. Beauftragung einer professionellen Suche z. B. bei DIMDI
4. Sonstiges

Bitte eintragen

(16) Hat Ihnen die Struktur unserer Antwort für Ihr eigenes Vorgehen bei Literaturrecherchen und Bewertung der Literatur geholfen bzw. verändert?

Bitte ankreuzen

1. Ja
2. Nein
3. Wenn nein, warum nicht?

.....
Bitte eintragen

(17) Welche kostenpflichtige online - Datenbank bzw. Literaturzugriffe haben bzw. nutzen Sie ?

(Mehrfachantworten möglich)

Bitte ankreuzen

1. Cochrane Library
2. ACP-Journal Club
3. Clinical Evidence
4. DIMDI
5. Andere Datenbanken

Bitte eintragen

Vielen Dank

EIDESSTATTLICHE VERSICHERUNG

gemäß Habilitationsordnung der Charité

Hiermit erkläre ich, dass

- keine staatsanwaltschaftlichen Ermittlungsverfahren gegen mich anhängig sind,
- weder früher noch gleichzeitig ein Habilitationsverfahren durchgeführt oder angemeldet wurde bzw. welchen Ausgang ein durchgeführtes Habilitationsverfahren hatte;
- die vorgelegte Habilitationsschrift ohne fremde Hilfe verfasst, die beschriebenen Ergebnisse selbst gewonnen wurden, sowie die verwendeten Hilfsmittel, die Zusammenarbeit mit anderen Wissenschaftlerinnen oder Wissenschaftlern und technischen Hilfskräften und die Literatur vollständig angegeben sind,
- mir die geltende Habilitationsordnung bekannt ist,

.....
Datum

.....
Unterschrift