EBM: Eine Einführung

Heike Arndt

Literatur – 12

Evidenzbasierte Medizin stellt in der Medizin in den letzten Jahren ein kräftig diskutiertes Schlagwort dar. Dabei ist festzustellen, dass oftmals die Definition, die Bedeutung, der Zweck, und die Inhalte von EBM sowie insbesondere der wissenschaftliche Kontext dem Anwender verschlossen geblieben sind. Nach David Sackett (1996), einem der aktuell bekanntesten und am häufigsten zitierten Protagonisten und Weiterentwickler von EBM wird das Konzept wie folgt definiert:

» EBM ist der gewissenhafte, ausdrückliche und vernünftige Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten. Die Praxis der EBM bedeutet die Integration individueller klinischer Expertise mit der bestmöglichen externen Evidenz aus systematischer Forschung.«

Diese Definition ist eine Zusammenfassung wissenschaftstheoretischer und wissenschaftsphilosophischer Überlegungen und birgt bei oberflächlicher Betrachtungsweise die Gefahr, als »false friend« zu falschen Rückschlüssen und zu einer falsch verstandenen, dadurch sogar vielleicht absurden und der eigentlichen Absicht von

EBM diametralen Anwendung zu führen, und man kommt nicht umhin, einzelne Begriffe zu definieren, um die wirkliche Aussage dieser Definition zu verstehen.

Weitere Schwierigkeiten beziehen sich offensichtlich auf die Anwendung der Instrumente von EBM. Allzu oft begegnet einem im Alltag, dass jemand seine Aussagen auf wenig Detailwissen über EBM bezieht, aber sich noch nicht einmal die Definitionen genauer angesehen hat, so dass durch die Anwendung der Instrumente (z. B. durch unreflektierten Rückgriff auf die Einteilung »Levels of Evidence«) sowohl die daraus erfolgenden Hypothesen, die eine Kausalität herleitenden Schlussfolgerungen, als auch die Anwendungen selbst im Sinne von EBM fehlerhaft sein können.

Hier können nur wenige Schlagworte wissenschaftstheoretischer und wissenschaftsphilosophischer Überlegungen angerissen werden. Anders als in früheren Zeiten – Jahrhunderten – beschäftigt man sich als Wissenschaftler heutzutage nicht mehr wie selbstverständlich mit Überlegungen darüber, was Wissenschaft, Wissen, Vernunft und Erkenntnis überhaupt ist. Sokrates, Platon, Aristoteles, Kopernikus, Galilei, Descartes, Hobbes, Locke, Spinoza, Kant, Leibniz, Goethe, Fried, Schweitzer, Hiatt, Einstein,

Weizsäcker, Bohr, Eccles, Popper – die Liste lässt sich durch viele Namen ergänzen.

Wissenschaft, Forschung und Philosophie gehörten untrennbar zusammen. Seit dem Zeitalter der Industrialisierung war es auch in der Medizin nicht mehr en vogue, diesen Kontext zu suchen. Die Wissenschaft – auch in der Medizin – hat sich den gleichen Spielregeln angepasst wie alle anderen Disziplinen der Wirtschaft und Industrie auch. Neue Erkenntnisse wurden in immer kürzeren Abständen generiert, oft mit inflationärer Qualität zugunsten einer Produktivität, die als Erfolgskriterium deklariert wurde.

Man kann von einer Wissensinszenierung sprechen, technokratisch fragmentiert, vom menschlichen wegführend und professionalisiert.

Hier bedeutet professionalisiert aber – wie schon dargestellt – eine Trennung von ethisch-moralischen Werten wie auch für die Wirtschaft beschrieben. Es geht hier um den spezifischen Leistungsvollzug überwiegend organbezogener Diagnosen und Therapien.

Erich Fromm beschrieb die gesellschaftlichen Entwicklungen folgendermaßen:

- » Die große Verheißung unbegrenzten Fortschritts – die Aussicht auf Unterwerfung der Natur und auf materiellen Überfluss, auf das größtmögliche Glück in größtmöglicher Zahl und auf uneingeschränkte persönliche Freiheit – das war es, was die Hoffnung und den Glauben von Generationen seit Beginn des Industriezeitalters aufrechterhielt.« (Aus: Erich Fromm: Haben oder Sein; DTV München, 2001, S. 13)

Produktivität und Systemwachstum im Sinne eines radikalen Hedonismus wurden in den Vordergrund gestellt. Dass hierdurch Schwierigkeiten entstanden, für die unsere Gesellschaft- in gesellschaftspolitischen sowie ethisch-moralischen Belangen – keine befriedigende Lösung gefunden hat, ist nahe liegend. Grundsätzliche, übergeordnete Überlegungen und Einstufungen wissenschaftlicher Erkenntnisse in ein gesellschaftliches, ethisch-moralisches Wertesystem haben in den letzten Jahrzehnten nur wenig stattgefunden.

Die Medizin ist eine Handlungswissenschaft, keine Grundlagenwissenschaft per se, und hält der Prüfung auf Wissenschaftlichkeit besonders dann Stand, wenn sie

- naturwissenschaftlich-biologisch fundiert ist,
- sich auf klinische Beobachtungen, Feststellungen und Messungen gründet,
- auf normativer, sozialer Grundhaltung basiert,
- und ihren therapeutischen Imperativ auf Zweckmäßigkeit prüft.

Die Art, mit der man seit Jahrtausenden medizinisches Wissen generiert hat, ist die Erfahrung, welche durch Lehre und Wirken erworben wurde. Medizin wurde bis vor kurzem als Heilkunst bzw. Heilkunde bezeichnet. Das wissenschaftliche Verständnis der Heilkunst wurde aus dem Bestreben der Experten erlangt, die Vorgänge, Ursachen und Wirkungsweisen der angewandten Methoden erklären zu wollen, die nach der Erfahrung zum Erfolg führten, oder es wurden auch Erklärungsansätze dafür gesucht, warum die gleiche Methode einmal funktionierte und heilbringend war, ein anderes Mal aber scheiterte (man denke beispielsweise an die Aderlasstherapie).

Aus wissenschaftstheoretischer Sicht stellte die Heilkunst somit eine nicht-analytische Wissenschaft dar. Charakteristisch für eine nicht-analytische Wissenschaft ist, dass sie ganzheitlich und individualisierend vorgeht. Sie macht sich die Fragen der Praxis zum Gegenstand ihrer Untersuchungen, um den Konflikt aufzulösen, den eine Praxis ohne Wissenschaft und eine Wissenschaft ohne Praxis mit sich bringen würde.

Der naturwissenschaftliche Entdeckungsprozess selbst wurde, wie oben beschrieben, fort-hin als ein konstruktiver Fabrikationsprozess verstanden. Gemäß wissenschaftstheoretischer Methoden wendete man sich von einem nicht-analytischen Ansatz hin zu einem analytischen Ansatz: »Eine analytische Vorgehensweise ist hiernach eine solche, die ihren Gegenstand in einzelne Bestandteile auflöst und die Beziehungen dieser Bestandteile untereinander betrachtet. In diesem Sinne »analytisch« arbeiten z. B. die Mathematiker, die formalen Logiker, die Sprach- »Analytiker«, die Naturwissenschaftler« (Aus:

Seiffert H: Einführung in die Wissenschaftstheorie 1; Becksche Reihe München 1996, S. 17).

» Die analytische Richtung hat den Versuch gemacht, das Problem der moralischen, sozialen oder politischen Wertung von Sachverhalten völlig aus der Wissenschaft herauszuwerfen, in dem man Wertfragen zu einer außerwissenschaftlichen, rein praktischen Angelegenheit erklärte.« (Seiffert, 1996, S. 21)

Historisch berühmtes Beispiel für die problematische Umsetzung einer solchen Vorgehensweise ist der Bau der Atombombe. In der Medizin betrifft dies beispielsweise Themen wie Transplantationen, Embryonenschutzgesetze, Gentechnikgesetze, Sterbehilfe, Abtreibungsgesetze.

Albert Schweitzer hat am 04.11.1954 in Oslo anlässlich der Entgegennahme des Friedensnobelpreises in seiner Rede mit dem Titel »Das Problem des Friedens in der heutigen Welt« diese Situation folgendermaßen beurteilt:

» Wagen wir die Dinge zu sehen, wie sie sind. Es hat sich ereignet, dass der Mensch ein Übermensch geworden ist ... Er bringt die übermenschliche Vernünftigkeit, die dem Besitz übermenschlicher Macht entsprechen soll, nicht auf ... Damit wird nun vollends offenbar, was man sich vorher nicht eingestehen wollte, dass der Übermensch mit dem Zunehmen seiner Macht zugleich immer mehr zum armseligen Menschen wird... Was uns aber eigentlich zu Bewusstsein kommen sollte und schon lange vorher hätte kommen sollen, ist dies, dass wir als Übermenschen Unmenschen geworden sind« (Hörkassette Nr. IV: Rede zum Friedensnobelpreis Oslo 1954, Albert-Schweitzer-Zentrum: <http://www.albert-schweitzer-zentrum.de/downl/art.pdf>).

Klar ersichtlich ist sein Standpunkt: Wissen ist Macht, und hier spiegelt sich die patriarchalische, oft dogmatische, evidenzbasierte Kultur der Medizinergesellschaft wider.

Wie schon dargestellt, zählt heutzutage mehr die Produktivität der Wissensgenerierung, und das Ziel ihrer Anwendung lag bisher kaum darin, wissenschaftliches Denken nach heutigem Verständnis zu spiegeln. Inwieweit bei dieser Wis-

sensinflation die Qualität des Erkenntnisprozesses und damit die Qualität der daraus resultierenden Information hochgehalten wird, ist sicherlich eine der Fragen, die sich aus diesem Kontext ergeben.

Der Pioniergeist und die Orientierung an ethisch-moralischen Werten, Visionen und Idealen sind wohl weitestgehend abhanden gekommen, die heutige, analytische Wissenschaft steht demnach oft im Dienste zur Erlangung lukrativer Posten und Finanzierungsmittel. Köbberling (Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis, DÄV Köln, 2000, S. 25) schreibt dazu:

» In der heutigen wissenschaftlichen Welt, ganz besonders im Umfeld der Medizin, ist der wichtigste Faktor bei der Unterdrückung von Zweifel (= wissenschaftliche Denkweise z. B. nach Popper, Anm. v. H.A) jedoch das Geld.«

Rienhoff (Qualitätsmanagement, aus: Das Public Health Buch, Hrsg. Schwartz FW, Badura B, Leidl R, Raspe H, Siegrist J. Urban und Schwarzenberg, 1998, S. 597) schreibt dazu:

» Gerade im universitären Bereich werden nach dem amerikanischen Vorbild der 60er Jahre die Regeln des »publish or perish« eingehalten – auch wenn die Evaluation wissenschaftlichen Arbeitens inzwischen nicht der Zahl, sondern der Qualität der Veröffentlichung den Vorzug gibt.« ... Da »allerdings erst nach längeren Perioden wissenschaftlich relevante Ergebnisse präsentiert werden können, verführt das universitäre Umfeld zu vielen Publikationen, die oft mehr den guten Willen als ein vorzeigbares Ergebnis dokumentieren.«

Zweifelhafte Qualitätskriterien wurden etabliert und angewendet, z. B. genügten lange Zeit alleine berufspolitische Machtpotenziale, um Arbeitsgruppen »erfolgreich« sein zu lassen.

Der Impactfaktor wird immer wieder diskutiert: Walder und Tramèr bemerken, »... dass die regelmäßige Publikation narrativer konventioneller Übersichtsartikel den »Impact Factor« des Journals positiv beeinflussen kann, da diese Artikel oft und kritiklos zitiert werden« (Evidence-based Medicine und systematische Reviews in

der perioperativen Medizin in: Der Anaesthetist, Springer 2001; 50 (9), S. 689–694). Man kann von einer Wissensinszenierung sprechen.

Welche Stilblüten so etwas treiben kann, ist in den einschlägig bekannt gewordenen Skandalen um medizinische Forschung in den letzten Jahren zu erkennen. Wissenschaftsbetrügereien werden zunehmend aufgedeckt, stigmatisiert und geahndet. Inwieweit hier eine Systemimmanenz zum Tragen kommt, kann an anderer Stelle diskutiert werden.

Jährlich erscheinen in ca. 20.000 biomedizinischen Zeitschriften mehr als 1 Mio. wissenschaftliche Arbeiten. Die Angaben über die Anzahl von randomisierten, kontrollierten Studien (RCT) sind so stark schwankend, dass man diese Zahlen am besten gar nicht zitiert, denn es wird schon aus diesen Angaben offensichtlich, wie unübersichtlich die Informationsflut ist. Laut Gerlach und Beyer (EBM in der Praxis des niedergelassenen Arztes aus: Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis, DÄV Köln, 2000, S. 69) müsste ein Arzt, wenn er alleine im Fach Innere Medizin einigermaßen up to date bleiben wollte, täglich wenigstens 17 Originalarbeiten studieren. Darüber hinaus müssten zahllose Artikel über neue diagnostische Tests, prognostische Faktoren, Epidemiologie und Pathophysiologie kritisch gelesen, beurteilt und in ihren Konsequenzen für die Praxis durchdacht werden. Dies ist für einen Durchschnittsmediziner – Menschen – nicht zu machen.

Das Resultat daraus ist zwangsläufig, dass es absolut unmöglich ist, überhaupt nur annähernd an alle (nur) allerwichtigsten Informationen zu kommen, schon gar nicht ist es möglich, diese hinsichtlich ihrer Vollständigkeit, Qualität, Relevanz, Übertragbarkeit usw. zu bewerten. Sprach- und Publikationsbiases beispielsweise sind bei dieser Betrachtung noch gar nicht berücksichtigt. Dennoch oder gerade deshalb:

! Die individuelle ärztliche Entscheidung braucht in der täglichen Anwendung das Korrektiv durch die Datenlage in der medizinischen Literatur.

Eine andere Facette kommt in der Definition von Sackett noch zum Ausdruck.

Wissenschaftsphilosophisch handelt es sich dabei um Überlegungen, was Wissen überhaupt ist und wie man zu Erkenntnissen gelangt. Rationalismus, dessen wichtigster Vertreter beispielsweise Descartes ist, geht davon aus, dass alle Erkenntnis nur auf dem Verstand basiert. Der Empirismus besagt, dass allein die Erfahrung die Quellen und Grenzen der Erkenntnis darstellt. Wichtige Vertreter sind Locke und Humes. Kant beispielsweise verknüpft die beiden Richtungen in seiner Auseinandersetzung »Die Kritik der reinen Vernunft«. Popper führt diese Theorien weiter aus, indem er – gemäß dem Zeitalter der Aufklärung – den Zweifel an einer Hypothese als wissenschaftliche Grundhaltung propagiert.

Nach Popper (Alles Leben ist Problemlösen, Piper 2002, S. 35) ist nur durch die Erfindung der kritischen Diskussion Wissenschaft erzeugbar. Alle naturwissenschaftlichen Theorien haben den Charakter eines vorläufigen Lösungsversuches, also einen hypothetischen Charakter. Wissenschaftlichkeit ergibt sich folglich daraus, den Zweifel an der aufgestellten Theorie gleich mitzuliefern, sozusagen: Eine Hypothese aufzustellen, die zur Falsifikation offen ist. Erst in dem Versuch, die Hypothese zu widerlegen, im sog. Eliminationsversuch, erlangt eine Hypothese die *Wahrscheinlichkeit* einer Wahrheit.

» Das besondere der Wissenschaft liegt in der bewussten Anwendung der kritischen Methode; Alle vorwissenschaftliche Erkenntnis, ob tierisch oder menschlich, ist dogmatisch; und mit der Erfindung der nicht-dogmatischen Methode, d. h. der kritischen Methode, beginnt die Wissenschaft.«

Wissenschaftliche Hypothesen können im Grundsatz nach 2 Prinzipien aufgestellt werden: dem der Deduktion und dem der Induktion. Die deduktive Methode der Hypothesenbildung beinhaltet den Beweis bzw. die Ableitung der Hypothese. Beweise können ausschließlich in der Mathematik und in der formalen Logik erbracht werden. Alle anderen Wissenschaften beruhen auf der Methodik der Induktion, angefangen von der Physik bis hin zur Medizin.

Die Methode der Induktion beinhaltet, dass man die Richtigkeit einer Hypothese, eines Sat-

zes, durch die Überprüfung von Einzelfällen in der Wirklichkeit feststellt. Typisch für diesen »induktiv-empirischen Weg ist die Erhebung und statistische Auswertung repräsentativer Stichproben (deskriptive Statistik) sowie die generalisierenden Rückschlüsse auf die Grundgesamtheit (Inferenz-Statistik)« (Eberhard, Einführung in die Erkenntnis- und Wissenschaftstheorie, 2. Aufl. 1999, Kohlhammer, S. 33). Diese Methode der Induktion ist die logische Basis der Wahrscheinlichkeitstheorie und schlussfolgernder Statistik.

Folglich können alle »lebenden« Wissenschaften nur versuchen, mit ihren Hypothesen der Wahrheit möglichst nahe zu kommen. »Die Idee der Annäherung an die Wahrheit setzt ... eine realistische Weltansicht voraus. Sie setzt nicht voraus, dass die Wirklichkeit so ist, wie sie unsere wissenschaftlichen Theorien beschreiben.« (Popper: Alles Leben ist Problemlösen, Piper 2002, S. 43).

Zu einer wissenschaftlichen Haltung gehören außerdem neben der Idee der größtmöglichen Annäherung an die Wahrheit und dem Zweifel an einer Hypothese auch die Objektivität des Wortes sowie eine unvoreingenommene Haltung.

Jede Hypothese hat nur den Charakter eines vorläufigen Lösungsversuches und kann durch eine neue Hypothese mit einer besseren Annäherung an die Wirklichkeit revolutioniert werden. Experimente können die Superiorität einer Theorie erweisen, aber nicht ihre Wahrheit. Siehe auch Thomas S. Kuhn »Die Struktur wissenschaftlicher Revolutionen«, Suhrkamp Taschenbuch Wissenschaft 1976, Frankfurt am Main.

Das heißt, jede Theorie zur Problemlösung muss sich einem Falsifikationsversuch unterziehen. Wenn sie zur Problemlösung beiträgt, dem Falsifikationsversuch also standhält, hat sie zur Lösung eines Problems, einer Fragestellung, beigetragen. Versuche, eine Theorie vor einer Falsifikation zu schützen, wird von Popper als »vorwissenschaftlich«, dogmatisch, betrachtet.

EBM ist ein bottom-up-Ansatz. Sie bietet theoretisch die Möglichkeit, nach einer Ära der Wissensgenerierung von unglaublichem Ausmaß das auf empirischem Wissen der *angewandten* Medizin basierende ärztliche Handeln mit einer analytischen Wissenschaftlichkeit (wieder) zu vereinigen.

Der Internist Bernhard Naunhyn schrieb 1905:

» Für mich ist es kein Zweifel, dass das Wort: 'Medizin wird eine Wissenschaft sein, oder sie wird nicht sein' auch für die Therapie gelten muss und gilt. Die Heilkunde wird eine Wissenschaft sein oder sie wird nicht sein! Mir ist sonnenklar, dass da, wo die Wissenschaft aufhört, nicht die Kunst anfängt, sondern rohe Empirie und das Handwerk« (Deutsche Revue 1905, 30; 343–355).

Weitere Ausführungen sprengen hier den Rahmen; es wird aber sicherlich deutlich, dass EBM unmittelbar im Kontext logischer wissenschaftstheoretischer und philosophischer Entwicklungen und Diskurse zu sehen ist und gar nicht den Anspruch hat, der Weisheit letzter Schluss zu sein. EBM propagiert sozusagen, erneuert und überholt zu werden. Somit komme ich zu dem Punkt der Definition von Evidence-based Medicine und ihrer Geschichte in Deutschland.

Eine Schwierigkeit besteht auch rein im Verständnis des Wortes »evidence«. Evidenz – ein Homonym: Im Lateinischen bedeutet Evidenz soviel wie augenscheinlich, einleuchtend, klar, offensichtlich. Bezogen auf einen griechischen Ursprung *enarges* heißt es soviel wie klar, anschaulich. Im Deutschen wie im Englischen wird Evidenz zwar per Lexika genauso definiert, im Sprachgebrauch aber gleichbedeutend gesetzt mit einem Tatbestand, der keines weiteren Beweises, keiner weiteren Überprüfung bedarf, also einer eigentlich auf juristische Belange ausgerichtete Bedeutungsübertragung.

Oft wird Beweis und Evidenz synonym verwendet. Im Kontext mit EBM bedeutet es aber soviel **wie wahrscheinlich(st)e Richtigkeit**, und diese Definition ist gemeint, wenn man von Evidence-based Medicine spricht: »available facts, circumstances, etc. indicating whether a thing is true or valid.« (Oxford English Dictionary, 1989, Oxford University Press).

Gemäß meiner vorangegangenen Ausführungen wird deutlich, dass es also in der EBM gar nicht um einen Beweis gehen kann. Das Wort »evidence« verkörpert eine nicht-analytische Wissenschaftsanschauung mit induktiv-empirischem Erkenntnisweg, wobei unter Anwendung

statistischer Methoden Fragestellungen mit Hypothesen, die immer mit Zweifel verbunden sind, mit einer größtmöglichen Annäherung an die Wahrheit beantwortet werden.

Diese wahrscheinlichste Richtigkeit beinhaltet die im Folgenden dargestellten Gütekriterien

- Objektivität, bzw. Verzerrungsfreiheit, d. h. Nachvollziehbarkeit von Methode und Ergebnis
- Reliabilität, d. h. die Zuverlässigkeit der Wiederholbarkeit
- Validität, d. h. die Methode misst genau das, was sie vorgibt zu messen, (sie bezieht sich auf verschiedene Aspekte, z. B. Inhalt, Konstrukt, Kriterien usw.) und auf
- Generalisierbarkeit, d. h. Übertragbarkeit der Ergebnisse in andere Umgebungen und Zusammenhänge, was auch als externe Validität bezeichnet wird

1994 veröffentlichten Hiatt und Goldman in *Nature* (371, S. 100) einen Artikel mit der Überschrift: »Making medicine more scientific«. Darin geht es *nicht* um den Ruf, die Grundlagenforschung auszubauen, sondern um die Evaluation *klinischer* Wissenschaft und Medizin.

EBM bezieht sich darauf, d. h. sie soll bei der medizinischen Entscheidungsfindung zur Versorgung des einzelnen Patienten dem Arzt ermöglichen, eine gewissenhafte, rückhaltlos kritisch geprüfte und v. a. zweckmäßige Anwendung seines Wissens auszuüben. EBM soll den ethischen Anspruch ärztlichen Verhaltens unterstützen und bietet eine gut lehrbare Methodik, um verantwortungsvolle Ärzte auszubilden.

Sie stellt eine Grundlage dafür dar, dass die ärztliche Berufsausübung nachvollziehbarer, weniger intuitiv und sicherer sein kann. Es kann keinesfalls ausreichen, Zweckmäßigkeit schlichtweg nur zu behaupten, schon gar nicht auf der Basis der grundsätzlich begrenzten und in aller Regel unkontrollierten eigenen klinischen Erfahrung.

Es ist gerade die individuelle Erfahrung und Kenntnis eines Arztes aus dem, was er gelernt hat,

und seinen klinischen Erfahrungen, die es möglich macht, externe Evidenz auf den einzelnen Patienten anzuwenden. Exzellente Forschungsergebnisse können für individuelle Patienten nicht anwendbar oder unpassend sein. Insofern muss ein Arzt entscheiden können, ob die externe Evidenz überhaupt in seine Handlung integriert werden kann. Entscheidungsprozesse werden vergleichsweise strukturierter, und die zielgerichtete Komponente medizinischer Entscheidungsfindung wird effektiver gestaltet.

Besonders hervorzuheben ist hier: Es geht somit um *die Integration von individueller klinischer Expertise mit der besten extern verfügbaren Evidenz*. Keines von beiden alleine ist eine hinreichende Basis für zweck- und patientenorientiertes Handeln, dem Grundsatz der EBM. In jedem Fall sollten wir nach der bestmöglichen externen Evidenz suchen, und falls diese nicht zu finden ist, die nächstbeste in Betracht ziehen.

Das Vorgehen in der Beantwortung einer klinischen Fragestellung beinhaltet folgende Schritte:

1. Zweifel am bisherigen Vorgehen, an der Gültigkeit von Lehrmeinungen in der Medizin, unsichere Kenntnisse bzw. Reflektion und ggf. Bestärkung des eigenen Wissens
2. Formulierung beantwortbarer Fragen: Dabei kommt der *richtigen* Fragestellung eine wichtige Bedeutung zu. Sie beinhaltet in der Regel 4 Aspekte:
  - Patientenorientiertes Problem
  - Frage nach einer Intervention
  - Frage nach einer Alternative
  - Frage nach einem relevanten Endpunkt
3. Suche nach der besten externen Evidenz (z. B. mittels Medline, Cochrane Library, Embase usw.)
4. Kritische Bewertung dieser Evidenz bezüglich Validität und klinischer Relevanz (kritische Bewertung s. unten Levels of Evidence, Consort Statement)
5. Umsetzung dieser Erkenntnisse in die klinische Arbeit
6. Bewertung der eigenen Leistung (Sackett berichtete, dass in der Medizinischen Klinik in Oxford bis zu 48% der spontanen Entscheidungen bei Visiten durch die Anwendung von EBM revidiert wurden.)

Mit einem sehr zentralen, häufig angeführten Missverständnis muss hier unbedingt nachdrücklich aufgeräumt werden:

EBM reduziert sich *nicht* auf randomisierte, kontrollierte Studien und Metaanalysen. Sie beinhaltet die Suche nach der *besten verfügbaren* externen Evidenz zur Beantwortung einer ganz konkreten, auf den individuellen Patienten ausgerichteten klinischen Fragestellung.

» Um etwas über die Genauigkeit eines diagnostischen Verfahrens zu erfahren, benötigt man gut durchgeführte Querschnittsstudien von Patienten, bei denen die gesuchte Krankheit klinisch vermutet wird – keine kontrollierte Studie. Für eine prognostische Fragestellung benötigen wir methodisch einwandfreie Follow-up-Studien von Patienten, die in einem einheitlichen, frühen Stadium ihrer Krankheit in die Studie aufgenommen wurden. Und manchmal finden wir die benötigte Evidenz in Grundlagendisziplinen wie Genetik oder Immunologie. Insbesondere bei der Frage nach Therapiemethoden sollten wir jedoch nicht-experimentelle Ansätze vermeiden, da diese häufig zu falsch-positiven Schlüssen hinsichtlich der Wirksamkeit von Maßnahmen kommen. Da randomisierte, kontrollierte klinische Studien und besonders systematische Übersichten dieser Studien uns mit höherer Wahrscheinlichkeit korrekt informieren und falsche Schlussfolgerungen weniger wahrscheinlich sind, wurden sie zum »golden standard« für die Beantwortung der Frage, ob Therapiemaßnahmen mehr nützen als schaden. Allerdings sind für manche Fragestellungen keine kontrollierten Studien notwendig (etwa erfolgreiche Interventionen bei sonst fatalen Konditionen), oder es bleibt keine Zeit für klinische Studien. Falls keine kontrollierte Studie für die besondere Situation unseres Patienten durchgeführt wurde, müssen wir die nächstbeste externe Evidenz finden und berücksichtigen« (aus Sackett: Was ist Evidenzbasierte Medizin und was nicht? In: Deutsches Netzwerk Evidenzbasierte Medizin e.V. 2002, http://www.ebm-netzwerk.de/was_ist_ebm.htm und: Sackett D, Rosenberg WMC, Gray JAM, Haynes, RB, Richards WS: Evidence based medicine: what it is and what it isn't. BMJ 1996; 312: 71–72).

Evidenzbasierte Medizin betrifft alle Bereiche der Medizin wie beispielsweise:

- Screening und Diagnostik
- Prävention und Rehabilitation
- Therapeutische Intervention
- Metaanalyse
- HTA Bericht
- Leitlinien
- Gesundheitsinformation für medizinische Laien
- Gesundheitsökonomische Studien

Als Messlatte dienen die Levels of Evidence, wobei *impliziert* wird, je höher eine Studie in dieser Hierarchie angeordnet wird, desto besser ist sie. Dies ist aber bei weitem nicht zwangsläufig der Fall.

Es genügt nicht, die Aussagen einer Studie anhand der Levels of Evidence in ihrer Qualität zu klassifizieren. Diese Rangordnung bezieht sich nur auf bestimmte Eigenschaften einer Studie, nämlich auf die **interne Validität** (nach Perleth M, Raspe H, 2001: Splitter 4 Levels of Evidence – Was sagen sie wirklich aus? Vorbemerkungen Deutsches Netzwerk EBM; <http://www.Netzwerk-ebm.de>).

Die interne Validität sagt nur etwas über den Grad der Freiheit systematischer Fehler aus.

Wie oben schon erwähnt, sagt ein statistischer Test beispielsweise nichts über die Richtigkeit der Versuchsplanung aus. Er untersucht lediglich die Frage, ob die Ergebnisse der untersuchten Stichprobe trotz der zufallsbedingten Streuung auf die Grundgesamtheit übertragbar sind, d. h. verallgemeinert werden dürfen (Induktionsschluss). Der Begriff »signifikant« macht auch keine Aussage darüber, ob das gefundene Ergebnis relevant ist bzw. ob der Versuchsaufbau korrekt ist. Das Ziel ist, systematische Fehler so gering wie möglich zu halten und zufällige Fehler zu minimieren. Die Beurteilung darüber gelingt also nicht nur anhand der bloßen Betrachtung statistischer Tests, was für sich genommen ja schon oft erhebliche Schwierigkeiten macht.

Wer nicht gerade ein ausgeprägtes mathematisches Talent hat, und das sind nach allgemeiner Einschätzung die wenigsten Menschen, hat mit der Anwendung und Interpretation von Statis-

tik alleine schon große Probleme. Die mühe-lose Bewertung von statistischen Testverfahren, von Fehlern verschiedener Ordnung und deren Bedeutung für die Praxis fällt wohl vielen schwer, auch denen, die wissenschaftlich tätig sind. Statistik ist eben ein Instrument, welches sich der intuitiven Einschätzung vieler Menschen hinsichtlich Aussage und Interpretation entzieht.

Selbst wenn eine Studie also formal den Ansprüchen RCT entspricht, muss sie noch lange nicht richtig oder gut sein. Eine gewissenhaft durchgeführte nicht randomisierte Studie kann, wenn die verzerrenden Einflüsse ausreichend berücksichtigt worden sind, im Ergebnis mit der eines RCT vergleichbar oder sogar überlegen sein.

Ein anderes Beispiel: Was nützt es, wenn nachgewiesen wird, dass ein nebensächlicher Laborparameter durch eine Therapie verändert werden kann, aber der Krankheitsverlauf und das Überleben eines Patienten nicht verbessert oder abgewendet werden können, im schlimmsten Fall nicht untersucht worden sind?

Also ist auch die Aussage über die Adäquanz, d. h. der klinische Angemessenheit, einer Studie für eine Bewertung wichtig. (■ Tabelle 1.1 und 1.2)

Auch die **externe Validität** einer Studie ist unbedingt mit in Betracht zu ziehen.

Hierunter versteht man die Fragestellung, ob die in der Studie gewonnenen Ergebnisse überhaupt übertragbar sind:

Die Möglichkeiten einer Anwendung oder Übertragbarkeit und damit Wertigkeit in dieser Hinsicht können bei einer bevölkerungsbasierten Beobachtungsstudie mit niedrigem Evidenzlevel deutlich höher sein als bei einer RCT.

Sehr problematisch wird die Beurteilung einer Studienlage auch dann, wenn eine Konsistenz von Studien zu einer Fragestellung zu untersuchen ist. Diese kann nur beurteilt werden, wenn **alle** verfügbaren Studien zu dieser Frage vorliegen. Was bringt es, wenn zwar mehrere RCT's, aber mit gegenteiligen Aussagen zur Verfügung stehen? Hier hilft die Bewertung anhand der Levels of Evidence ebenfalls nicht weiter.

■ Tabelle 1.1. Level of Evidence

Grade of Recommendation	Level of Evidence	Therapy/Prevention, Aetiology/Harm	Prognosis	Diagnosis	Economic analysis
A	1a	SR (with homogeneity) of RCTs	SR (with homogeneity) of inception cohort studies; or a CPG validated on a test set.	SR (with homogeneity) of Level 1 diagnostic studies; or a CPG† validated on a test set.	SR (with homogeneity) of Level 1 economic studies
	1b	Individual RCT (with narrow confidence intervall)	Individual inception cohort study with >80% follow-up	Independent blind comparison of patients from an appropriate spectrum of patients, all of whom have undergone both the diagnostic test and the reference standard	Analysis comparing all (critically-validated) alternative outcomes against appropriate cost measurement, and including a sensitivity analysis incorporating clinically sensible variations in important variables
	1c	All or none	All or none case-series	Absolute SpPins and SnNouts	Clearly as good or better, but cheaper. Clearly as bad or worse but more expensive. Clearly better or worse at the same cost.

Tabelle 1.1. Fortsetzung

Grade of Recommendation	Level of Evidence	Therapy/Prevention, Aetiology/Harm	Prognosis	Diagnosis	Economic analysis
B	2a	SR (with homogeneity) of cohort studies	SR (with homogeneity) of either retrospective cohort studies or untreated control groups in RCTs.	SR (with homogeneity) of Level >2 diagnostic studies	SR (with homogeneity*) of Level >2 economic studies
	2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Retrospective cohort study or follow-up of untreated control patients in an RCT; or CPG not validated in a test set.	Any of: 1. Independent blind or objective comparison; 2 Study performed in a set of non-consecutive patients, or confined to a narrow spectrum of study individuals (or both) all of whom have undergone both the diagnostic test and the reference standard; 3 A diagnostic CPG not validated in a test set.	Analysis comparing a limited number of alternative outcomes against appropriate cost measurement, and including a sensitivity analysis incorporating clinically sensible variations in important variables.
	2c	»Outcomes« Research	»Outcomes« Research		
	3a	SR (with homogeneity*) of case-control studies			
	3b	Individual Case-Control Study		Independent blind or objective comparison of an appropriate spectrum, but the reference standard was not applied to all study patients.	Analysis without accurate cost measurement, but including a sensitivity analysis incorporating clinically sensible variations in important variables.
C	4	Case-series (and poor quality cohort and case-control studies §§)	Case-series (and poor quality prognostic cohort studies)	Any of: 1. Reference standard was unobjective, unblinded or not independent; 2. Positive and negative tests were verified using separate reference standards; 3. Study was performed in an inappropriate spectrum of patients.	Analysis with no sensitivity analysis
D	5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or »first principles«	Expert opinion without explicit critical appraisal, or based on physiology, bench research or »first principles«	Expert opinion without explicit critical appraisal, or based on physiology, bench research or »first principles«	Expert opinion without explicit critical appraisal, or based on economic theory

These levels were generated in a series of iterations among members of the NHS R&D Centre for Evidence-Based Medicine (Chris Ball, Dave Sackett, Bob Phillips) <http://www.cebm.net>

■ **Tabelle 1.2.** Levels of Evidence für Therapie- und Präventionsstudien

Grad	Studiendesign
Ia	Metaanalysen: Systematische Übersicht von randomisierten, kontrollierten Studien (mit Homogenität)
Ib	Einzelne randomisierte, kontrollierte Studien mit engem Konfidenzintervall
IIa	Systematische Übersicht aus Kohortenstudien (mit Homogenität)
IIb	Einzelne Kohortenstudien/randomisierte, kontrollierte Studie mit methodischen Mängeln
IIc	Outcome-Forschung
IIIa	Systematische Übersicht aus Fall-Kontroll-Studien
IIIb	Einzelne Fall-Kontroll-Studie
IV	Fallserien, Kohortenstudien, Fall-Kontroll-Studien mit methodischen Mängeln
V	Konsensuskonferenzen, Erfahrungen von anerkannten Autoritäten ohne explizite Grundlage von kritisch bewerteter Evidenz

Ein weiteres Problem stellen Publikationsbiases und Sprachbiases in der Betrachtung von Studien dar. Die Beurteilung der Effektivität einer therapeutischen Intervention ist ebenfalls nicht am Level of Evidence festzumachen. Hier benötigt man zur Abschätzung die »numbers-needed-to-treat« (NNT), d. h. die Anzahl der Patienten, die behandelt werden müssen, um ein negatives bzw. zusätzlich schädigendes Ereignis zu vermeiden.

Fragen zur Beurteilung der Wirksamkeit (efficacy) eines Studienergebnisses und nach dem Nutzen (effectiveness) eines solchen müssen ebenfalls unter anderen Gesichtspunkten beantwortet werden. Es geht hier im Wesentlichen um die Beurteilung einer Übertragbarkeit von Ergebnissen aus den Idealbedingungen kontrollierter Studien (efficacy) in den medizinischen Alltag und den Nutzen, also die Wirksamkeit unter Alltagsbedingungen (effectiveness).

Zahlreiche weitere Maßzahlen werden herangezogen, um zu beurteilen, welche Stärke die Effekte haben, d. h. um wie viel sie uns unserem Ziel näher bringen oder um Belege dafür zu erhalten, dass es eben **diese** Interventionen sind, die

uns dem Ziel wesentlich näher gebracht haben. Gemeint ist, dass der Effekt nicht zufällig, durch Verzerrungen (biases) oder durch Mitursachen (confoundings), hervorgebracht wurde.

Beispiele für weitere Maßzahlen:

NNT	Numbers needed to treat 1/ARR
NNH	Numbers needed to harm
CER	Control event rate/Ereignis in der Kontrollgruppe
EER	Experimental event rate/Ereignisse in der Behandlungsgruppe
RRR	Relative Risikoreduktion CER-EER/CER
ARR	CER-EER
Odds-Ratio	
Konfidenzintervalle	

Es ist – um es nochmals ganz deutlich auszudrücken – hinsichtlich bestimmter Fragestellungen

in der Medizin – schlichtweg unsinnig anzunehmen, dass nur das Gültigkeit hat, was durch RCT untersucht und für signifikant erklärt worden ist.

Da die beschriebenen Levels of Evidence und die bloße Statistik zur Bewertung von Studien sowie die Kriterien zur Beurteilung der externen Validität noch nicht ausreichen, sucht man nach weiteren Beurteilungskatalogen, die es ermöglichen, die richtigen Fragen an die Studien zu stellen und **v. a. damit** die Wertigkeit zu beurteilen.

Einer dieser Beurteilungskataloge ist das zurzeit wohl am weitesten verbreitete CONSORT Statement: Consolidated Standards of Reporting Trials hilft, randomisierte, kontrollierte Studien zu bewerten (Ohrer D, Schulz KF, Altman DG: The Consort Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials. *Ann. Int. Med.* 2001, 134 (8): 657-662; <http://www.annals.org> und Falck-Ytter Y, Blümle A, Schulgen G, Antes G, Deutsches Cochrane Zentrum »Das CONSORT-Statement: Verbesserung der Berichterstattung über randomisierte klinische Studien Deutsches Cochrane Zentrum, 2002.« Im Jahre 1996 wurde eine Leitlinie für die Publikation klinischer Therapiestudien, das CONSORT-Statement, verfasst (Begg et al., 1996).

Inzwischen ist diese Leitlinie Teil der Autorensrichtlinien von über 70 internationalen Fachzeitschriften. Eine aktualisierte Fassung des CONSORT-Statements mit einer ausführlichen Beschreibung der einzelnen Qualitätskriterien wurde im Jahre 2001 publiziert (Moher et al., 2001; Altman et al., 2001). Das CONSORT-Statement gibt Hilfestellungen für Autoren bei der Publikation einer klinischen Studie in Form einer Checkliste und eines Flussdiagramms und soll zur Verbesserung der Publikationen über klinische Studien beitragen. Weitere Infos sind beispielsweise unter <http://www.consort-statement.org> zu erhalten.

EBM wurde weiterentwickelt und stellt mit den Methoden der Epidemiologie und Instrumenten der Statistik und Bewertungsmaßstäben Möglichkeiten zur Verfügung, in dem Dschungel der Medizin wissenschaftliche Kriterien aufzuzeigen, die es dem einzelnen Arzt ermöglichen sol-

len, die Flut des Erkenntnisgewinns systematisch zu sichten und gemäß der dargestellten Intentionen zu bewerten und anzuwenden.

Gerd Antes vom Cochrane-Zentrum in Freiburg, deutscher Verfechter der EBM, verweist im Rahmen politischer Diskussionen immer wieder streng auf die eigentliche Intention und Technik der EBM. Es beklagt eben auch, dass EBM von außen häufig in ihrer Intention und Komplexität reduziert wird auf wenige Aspekte, um damit in der politischen Diskussion Macht auszuüben. In seinem Vorwort zum Jahresbericht 2002 des Deutschen Netzwerkes EBM schreibt er: »Nicht zu übersehen ist, dass der Eingang von EBM in die politische Diskussion und der damit verbundene Missbrauch einerseits sowie die Realität der ärztlichen Praxis andererseits erschweren, die Ärzteschaft zu erreichen.« (http://www.ebm-netzwerk.de/jahresbericht_dnebm_2002.pdf)

Im Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis (DÄV Köln, 2000) schreibt Raspe H (S. 42): »EbM ist Ausdruck und Instrument einer Rationalisierung der Medizin; Rationalisierung in dem Doppelsinn, »das Vernünftige sparsam tun.«

Im SGB V § 2 Leistungen wird darauf Bezug genommen:

- » (1) Die Krankenkassen stellen den Versicherten die im Dritten Kapitel genannten Leistungen unter Beachtung des Wirtschaftlichkeitsgebots (§ 12) zur Verfügung, soweit diese Leistungen nicht der Eigenverantwortung der Versicherten zugerechnet werden. Behandlungsmethoden, Arznei- und Heilmittel der besonderen Therapierichtungen sind nicht ausgeschlossen. Qualität und Wirksamkeit der Leistungen haben dem allgemein anerkannten Stand der medizinischen Erkenntnisse zu entsprechen und den medizinischen Fortschritt zu berücksichtigen.«...
- »(4) Krankenkassen, Leistungserbringer und Versicherte haben darauf zu achten, dass die Leistungen wirksam und wirtschaftlich erbracht und nur im notwendigen Umfang in Anspruch genommen werden.«

Rationalisierung im Sinne EBM meint aber v. a. nicht (nur) Wirtschaftlichkeit, sondern die Er-

schließung von Vernunftreserven, und zwar: Wissenschaftlichkeit.

Nochmals: EBM soll Medizin mehr wissenschaftlich machen – d. h. im Sinne Poppers eben eine vernünftige Argumentationsbasis herbeizuführen: Zweifel, Hypothese, Falsifikation.

EBM schützt sich selbst nicht vor Missbrauch durch falsch verstandene, unvollständige oder unkritische Erfassung durch den Anwender, oder auch nicht davor, seit Jahren mit haarsträubenden Argumenten entweder in der Luft zerrissen oder enthusiastisch gefeiert zu werden. Sie schützt sich auch nicht davor, dass man, bei Anwendung der Methode von EBM auf EBM selbst nicht immer zu einem Konsens findet. Denn:

! »Die Praxis der EBM bedeutet die Integration individueller klinischer Expertise mit der bestmöglichen externen Evidenz aus systematischer Forschung.«

Wenn trotz ausreichender Definition und Darstellung des wissenschaftstheoretischen Kontextes keine interne Evidenz für EBM vorhanden ist, kann sie nicht im eigenen Sinne richtig verstanden und angewendet werden. Vielleicht liegt in diesem daraus entstehenden Paradoxon der einzig mögliche Kritikpunkt an EBM. Technologien per se genommen sind immer neutral. Ob sie zum Segen oder Fluch werden, hängt von den Wertvorstellungen der Menschen ab.

! Belassen wir die Theorie und Philosophie von EBM doch ganz einfach dort, wo sie sich selbst sieht: Als eine an ein Wertesystem geknüpfte Betrachtungsweise und Methode zum wissenschaftlichen Diskurs, möglicherweise dazu geeignet, den eingangs beschriebenen Konflikt aufzulösen, den eine Praxis ohne Wissenschaft und eine Wissenschaft ohne Praxis mit sich bringen würde. Also: Eine Verknüpfung von analytischer und nicht-analytischer medizinischer Wissenschaft unter Berücksichtigung unseres heutigen gesellschaftlichen Wertesystems und konjunkturellen Ausrichtung. Sie kann Ausgangspunkt für eine Neuausrichtung einer Professionalisierung in der Medizin in Klinik und Wissenschaft darstellen.

Somit ist es an uns, dem Wissenschaftler, professionell mit EBM umzugehen und damit zu verhindern, dass einzelne Teile für Zwecke benutzt werden, sei es in der Gesundheitssystemgestaltung, im Gesundheitsmanagement, in der Gesundheitsökonomie, in der Politik, Rechtsprechung usw., die eigentlich gar nicht auf EBM bezogen werden können.

» Wer will was Lebendiges erkennen und beschreiben, sucht erst den Geist herauszutreiben. Dann hat er die Teile in der Hand – fehlt leider nur das geistige Band. (J. W. v. Goethe)

Literatur

Oxford Center of EbM: <http://www.cebm.net/index.asp>

Deutsches Netzwerk EBM: <http://www.ebm-netzwerk.de>

Cochrane: <http://www.cochrane.de>

Sackett, D.L., Richardson, W.S., Rosenberg, W., Haynes, R.B.: Evidenzbasierte Medizin Zuckschwerdt-Verlag München, 1999

Kunz R, Ollenschläger G, Raspe H, Jonitz G, Kolkman F (Hrsg.): Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis Deutscher Ärzte-Verlag 2000