

NORBERT DONNER-BANZHOF

Was zählt als Argument? Belege und Begründungen bei medizinischen Entscheidungen

Abstract

Entscheidungen in Medizin und Gesundheitsversorgung sind ohne einen Bezug zur Evidenzbasierten Medizin (EbM¹) kaum noch denkbar. Dieser Begriff steht für einen sorgfältigen, ausgewogenen und ausdrücklichen Bezug auf einschlägige klinische Studien (Sackett et al. 1996).

Geht es um die Rechtfertigung medizinischer Maßnahmen, ist die Unterscheidung von zwei Arten von Entscheidungen zentral: der a) Repertoire- und der b) klinischen Entscheidung. Die erstere trifft eine Behörde bei der Marktzulassung eines Medikaments oder ein Gesundheitssystem in Bezug auf die Kostenübernahme einer Technologie. Aber auch Praxen und Krankenhäuser müssen ihr Repertoire an diagnostischen und therapeutischen Maßnahmen festlegen. Dieses stellt für die klinische Entscheidung an der einzelnen Patientin nur den Hintergrund dar. Es sind die Argumentationen über Repertoire-Entscheidungen, welche sich durch die Etablierung der EbM grundlegend verändert haben.

Die Etablierung der EbM hat zu Machtverschiebungen geführt. Da Studien, welche den Kriterien der EbM genügen, sehr teuer sind, hat dies die Rolle von pharmazeutischer und Medizinprodukt-Industrie gestärkt und zu einer Oligopolisierung geführt. Traditionelles persönliches Wissen hat heute gegenüber digitalen Datenbasen, Zeitschriften und Lehrbüchern eine geringere Bedeutung. Studien im Sinne der EbM sind Voraussetzung für die Entwicklung von Entscheidungshilfen, mit welchen sich Patientinnen an Entscheidungen über ihre Behandlung beteiligen können. Vor dem Hintergrund kontroverser Diskussion um die EbM ist vor allem die weitgehende Akzeptanz der EbM-Standards bemerkenswert.

For several decades, Evidence-based Medicine (EBM) has become the most important frame of reference for decisions in health care. It denotes "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett et al. 1996).

¹ Während im Englischen die Abkürzung „EBM“ lautet, wird in Deutschland „EbM“ vorgezogen, um eine Verwechslung mit der Gebührenordnung zu vermeiden, die für Leistungen in der ambulanten Versorgung gilt (Einheitlicher Bewertungsmaßstab).

Good patient care requires decisions at two levels: a) the medical repertoire; b) the individual clinical decision. The former must be made by an authority granting market access or by the health care system regarding coverage. Hospital departments, practices and individual clinicians must also decide which components should be part of their repertoire. For the individual clinical decision, the repertoire of available treatment options provides only the background. It is for the repertoire decision for which EBM has fundamentally changed what counts as an argument and what does not.

The global spread of EBM has resulted in shifts of power and authority. Since trials meeting the methodological requirements of EBM are very expensive, the pharmaceutical and device industry has become a more powerful player in an oligopolistic market. Traditional personal academic knowledge and influence have been superseded by impersonal sources of knowledge, such as electronic databases, journals, and textbooks. Studies as favoured by EBM are necessary to develop decision aids aiming at involving patients for decisions affecting their health care. That the standards of EBM are almost universally accepted worldwide is most remarkable.

Keywords: Evidenzbasierte Medizin, klinische Entscheidung, System-Entscheidung, regulatorische Entscheidung, Argumentation

Evidence-based Medicine; clinical decision, policy-decision, regulatory decision, argumentation

1 Hintergrund

Vor etwa 40 Jahren tauchten Begriffe wie *Outcome Research* und *Evidence-based Medicine* (EbM) in der Diskussion um die Gesundheitsversorgung und die Medizin auf. Sie standen damals für eine neuartige Begründung und Rechtfertigung von präventiven, diagnostischen und therapeutischen Maßnahmen.

Ich werde zunächst einen kurzen historischen Abriss über das Auftreten der Bezeichnung EbM geben, der auch den sozialen und gesundheitspolitischen Kontext einschließt (Abschn. 1.1 und Abschn. 1.2). Sodann definiere ich zwei Arten von Entscheidungen, für welche die EbM in unterschiedlichem Maße relevant ist (Abschn. 2). Die zentralen Entscheidungsgremien des Gesundheitswesens (z. B. Gemeinsamer Bundesausschuss [G-BA], Leitlinien-Entwicklung) sind explizit auf die Maßstäbe der EbM festgelegt (Abschn. 3). Ich werde die dabei „zulässigen“ Argumente darstellen (Abschn. 4), aber auch die Alternativen, welche durch die EbM verdrängt bzw. relativiert worden sind (Abschn. 5). Obwohl die EbM primär der Rezeption wissenschaftlicher Befunde dient, hat sie Rückwirkungen auf die Entwicklung und Verbreitung von Entscheidungshilfen, die klinische Forschung und die Machtverhältnisse im Gesundheitswesen, den Universitäten und dem Zugang zu Wissen gehabt (Abschn.6). Schließlich gehe ich auf Kontroversen in Zusammenhang mit der EbM ein (Abschn. 7), obwohl es vor allem die universelle Akzeptanz der EbM ist, die herauszustellen ist (Abschn. 10).

1.1 Evidenzbasierte Medizin

Im Jahr 1981 erschien im kanadischen Ärzteblatt (CMAJ) die erste Folge einer Serie „*How to read clinical journals*“ (Department of Clinical Epidemiology and Biostatistics, McMaster University 1981), verfasst von Angehörigen des Department of Clinical Epidemiology and Biostatistics der McMaster University in Hamilton (Ontario, Kanada). Sie zielte auf vielbeschäftigte Ärztinnen², die mit ihrer Fachlektüre nicht mehr nachkommen. Ihnen wurde ein Algorithmus an die Hand gegeben, mit welchem sie Publikationen ohne Relevanz und mit unzureichender Methodik frühzeitig aussortieren und damit ihre Zeit effizienter einsetzen konnten. Dies lief vor allem darauf hinaus, Studien zu Mechanismen (meist Tierversuche) oder Meinungsäußerungen zu ignorieren und sich stattdessen auf Originalarbeiten mit valider Methodik und unmittelbarer klinischer Konsequenz zu konzentrieren. Dies entspricht den „klinischen Studien“ im engeren Sinn, als zuständiges Fach spricht man auch von „Klinischer Epidemiologie“.

Zwölf Jahre später startete dieselbe Gruppe u. a. mit David Sackett, Gordon Guyatt und Andy Oxman die „*Users' Guides to the Medical Literature*“ im US-amerikanischen Ärzteblatt (JAMA) (Oxman et al. 1993), eine Serie, die auch als Buch veröffentlicht wurde (Guyatt 2002) und heute noch weitergeführt wird. Hier wird die ‚Evidence-based Medicine‘ als Begriff eingeführt und die „EbM-Sequenz“ stärker betont, d. h. aus der eigenen praktischen Arbeit eine Frage zu formulieren, Recherchen anzustellen, eine passende Original-Publikation zu finden, diese methodisch zu würdigen und ggf. eine Konsequenz für die Behandlung der eigenen Patientinnen zu formulieren.

In den deutschsprachigen Ländern gründeten 1998 Angehörige von Universitäten, Fachgesellschaften, Institutionen des Gesundheitswesens und Gesundheitsberufen in Praxis und Krankenhaus eine Fachgesellschaft zur Durchsetzung der EbM-Perspektive: das *Netzwerk Evidenzbasierte Medizin*. Seit dem Jahr 2000 nimmt das *Sozialgesetzbuch V* (SGB V) ausdrücklich auf „evidenzbasierte“ Leitlinien Bezug.

In Bezug auf Struktur und Inhalte haben die o. g. Serien viele Zeitschriften, Kolumnen, Lehrbücher und Kurse inspiriert. Bald nach der Jahrtausendwende wurde die Anwendung der EbM institutionalisiert. Das britische National Institute for Health and Care Excellence (NICE), die US-amerikanische Agency for Healthcare Research and Quality (AHRQ) und in Deutschland das von Bundesärztekammer und Kassenärztlicher Bundesvereinigung getragene Ärztliche Zentrum für Qualität in der Medizin (ÄZQ) sind bekannte Beispiele. Einen gesetzlichen Auftrag hat in Deutschland das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Dieses und das NICE verwenden explizit die Standards der EbM, um für das Gesundheitssystem die Aufnahme neuer Behandlungsverfahren zu evaluieren. Das britische NICE trifft diese Entscheidung selbst, während das deutsche IQWiG den eigentlich entscheidenden G-BA berät.

2 Mit der weiblichen Form sind in der Regel auch die übrigen Geschlechter gemeint.

1.2 Historischer und gesellschaftlicher Kontext

Im traditionellen Verständnis professioneller Rollen war die explizite Rechtfertigung medizinischen Handelns viel weniger kritisch als heute. Professionen hatten das Privileg, eigene Normen zu entwickeln und deren Einhaltung zu überwachen. Staatliche Eingriffe beschränkten sich auf die Definition des Berufsbildes einschließlich der Zugangsvoraussetzungen (Studium, Prüfungen) und den Auftrag an eine zuständige Organisation (Kammern, College usw.). In Bezug auf Regelwerke ihres Handelns (Berufsordnungen) hatten die Professionen selbst einen großen Gestaltungsspielraum (Freidson 1995).

Staatliche, staatlich regulierte und private Kostenträger (Versicherungen) wurden im Laufe des 20. Jahrhunderts zu wichtigen Akteuren der Daseinsfürsorge. Während das Krankengeld (Ersatz des Lohnausfalls durch Krankheit) die wichtigste Leistung einer Krankenkasse im späten 19. Jahrhundert war, wuchsen mit der Zahl medizinischer Maßnahmen zwar die Möglichkeiten von Diagnose, Therapie und Prävention, aber auch das Kostenrisiko von Erkrankungen und Unfällen. Krankenkassen und staatliche Institutionen sorgten deshalb dafür, dass das Vertrauen in das Handeln von Professionellen zunehmend von Kontroll- und spezifischen Rechtfertigungsmechanismen abgelöst wurde. Diese bezogen sich auf prospektive Einzelfallentscheidungen (z. B. Genehmigung der Krankenkasse als Voraussetzung für eine Behandlung), retrospektive Analyse von Verwaltungsdaten (z. B. Wirtschaftlichkeit von Krankenhausbehandlungen, Verordnungen von Medikamenten und Heilmitteln), allgemeine Entscheidungen über die Zulässigkeit (Kostenübernahme) einer Methode u. v. m. Untersuchungen über nicht gerechtfertigte Variationen in der Versorgung hatten kritische Fragen provoziert.

Wenn die Häufigkeit von Operationen regional fünf- oder siebenfach variierten (Wennberg/Gittelsohn 1982), ohne dass sich dies demografisch oder epidemiologisch erklären ließ, war der Vorwurf professioneller Willkür nicht weit. Mit einer Orientierung an einschlägigen wissenschaftlichen Untersuchungen und deren Aufbereitung, wie z. B. Leitlinien, sollte die gesundheitliche Versorgung rationaler, wirksamer, sicherer und einheitlicher werden. Dieser Anspruch hat die Gesundheitssysteme tiefgreifend verändert. Bei der Bewertung und Akzeptanz von Argumenten hat sich diese Veränderung am deutlichsten gezeigt.

1.3 Repertoire- versus klinische Entscheidung

In einem differenzierten, sich an wissenschaftlichen Ergebnissen orientierenden Gesundheitssystem sind grundsätzlich auf zwei Ebenen Entscheidungen zu treffen (Donner-Banzhoff 2022, Kap 5.2): zum einen die klinische, zum anderen die Repertoire-Entscheidung. Letztere bezieht sich darauf, ob eine Maßnahme in ein Repertoire aufgenommen wird. Dies mögen die in einem Land (Staat) verordnungsfähigen Medikamente sein, die von Kostenträgern (Krankenkassen) bezahlten operativen Verfahren, oder die Empfehlungen einer landesweit gültigen Leitlinie zum diagnostischen Vorgehen bei Brustschmerzen.

Eine Repertoire-Entscheidung trifft aber auch eine Krankenhaus-Abteilung, wenn sie Abläufe für häufige Behandlungsprobleme schriftlich festlegt. Sogar einzelne Ärztinnen müssen solche Entscheidungen treffen; so gilt es als ein Merkmal guter Medikamenten-Verordnung, wenn aus einer möglicherweise großen Auswahl nur wenige Wirkstoffe eingesetzt werden, deren Wirkdauer, Nebenwirkungsprofil, Interaktionspotenzial mit anderen Medikamenten usw. der ärztlichen Verordnerin gut bekannt sind.

Die klinische Entscheidung in Bezug auf Diagnose und Therapie dagegen wird im Angesicht der einzelnen Patientinnen getroffen. Grundlage des Vorgehens ist das o. g. Repertoire; aber jetzt kommen zu den allgemeinen Kriterien die individuellen hinzu: Bestehen weitere Erkrankungen, welche bei der medikamentösen Behandlung bedacht werden müssen? Sind Allergien bekannt? Lehnen die Patientinnen Spritzen ab? – um nur einige Beispiele zu nennen. Die Repertoire-Entscheidung wird unabhängig von individuellen Patientinnen auf Grund von allgemeinen Einsichten und an Populationen gewonnenen empirischen Befunden getroffen; diese treten dagegen in den Hintergrund bei der klinischen Entscheidung, wo die Charakteristika und Präferenzen der Patientin im Vordergrund stehen. Bei der klinischen Entscheidung ist ein Abweichen von allgemeinen Empfehlungen häufig und durch die Charakteristika des Einzelfalls zu rechtfertigen. Aus der Behandlung mag sich eine auf das Repertoire bzw. die allgemeinen Wissensbestände der Medizin zielende Frage ergeben, die der oben erwähnten EbM-Sequenz entspricht. Allerdings werden Antworten darauf (immer noch) überwiegend durch Befragung von Kolleginnen, Recherche in hochaufbereiteten Datenbanken einschließlich Leitlinien, Überweisung der Patientin, nur selten durch Rekurs auf eine wissenschaftliche Originalarbeit beantwortet (Ebell 2009).

Mit der Unterscheidung der beiden Bereiche ist gleichzeitig der Geltungsbereich der EbM angesprochen: ihre bevorzugte Domäne ist die Repertoire-Entscheidung. Hier hat sie für eine dramatische Veränderung von Strukturen und Standards der Argumentation gesorgt, während eine ethnologische Beobachterin in Praxen und Krankenhäusern heute deutlich geringere Unterschiede zum 20. Jahrhundert bemerken würde. Letztere umfassen viele z. T. sehr alte Traditionsbestände, die sich nur langsam verändern – auch wenn sich die eingesetzten Technologien³ über die Jahrzehnte natürlich verändert haben (Donner-Banzhoff 2022, Kap. 12). Die Unterscheidung von Repertoire- und klinischer Entscheidung kann m. E. zahlreiche Kontroversen über die (Nicht-) Anwendung und (Nicht-) Rechtfertigung der EbM entschärfen, wenn nicht gar als Missverständnis auflösen. Ich werde mich in dieser Arbeit vorzugsweise auf die Repertoire-Entscheidung beziehen, dabei allerdings die große Spannweite der Kontexte im Auge haben, die vom G-BA (Kostenübernahme für ein neues Verfahren) bis zum informellen kollegialen Gespräch über bessere und schlechtere Behandlungen reicht.

³ Ich verwende diesen Begriff sehr allgemein für Interventionen jenseits unmittelbar persönlicher Effekte; Medikamente sind hier also miteingeschlossen

2 Orte der Argumentation

Die Notwendigkeit einer Argumentation ergibt sich dann, wenn etwas strittig, widersprüchlich oder defizitär erscheint. Dies mag die erhöhte Nebenwirkungsrate eines Medikaments sein (soll es aus der Medikamentenliste des Krankenhauses gestrichen werden? Oder gar seine Marktzulassung verlieren?) Oder behauptet eine pharmazeutische Firma die Überlegenheit eines neuen Medikaments im Vergleich zu den vorhandenen Behandlungsoptionen, sog. ‚Claim‘⁴? Im Kontext der Gesundheitsversorgung müssen diese Fragen entschieden werden (Streichung oder unveränderter Status des Medikaments; Kostenübernahme⁵ des neuen Medikaments JA oder NEIN). Die resultierende Antwort beansprucht Geltung für den jeweiligen Bereich.

Repertoire-Entscheidungen können unterschiedliche Konfliktpotenziale beinhalten. Die Gruppe der Oberärztinnen einer Krankenhaus-Abteilung, die das Vorgehen bei häufigen Notfällen festlegen, hat einen gemeinsamen fachlichen Hintergrund und professionelle Interessen. Sie können sich auf wissenschaftliche Belege, Praktikabilität und Kosten als Kriterien für interne Regelungen konzentrieren. Diese Situation ist von einem geringen Grad an „Agonalität“ gekennzeichnet (Hannken-Illjes 2018, 26). Dem entgegengesetzt ist die Situation im G-BA, dessen Mitglieder („Bänke“) einen Auftrag als Interessensvertreterinnen haben. Dabei steht u. U. viel auf dem Spiel: eine neue Methode mag einer Fachgruppe von Ärztinnen und/oder einem Unternehmen neue Umsatzmöglichkeiten eröffnen, andere jedoch in ihrer Existenz gefährden. Die gesetzlichen Krankenkassen müssen hohe Ausgaben vermeiden, da sie durch höhere Beiträge zahlende Mitglieder an konkurrierende Kassen verlieren können. Interessenskonflikte sind also die Regel, eine rationale, ausschließlich an der Sache orientierte Argumentation stellt hier, wie auch in Leitliniengruppen, eher eine Ausnahme dar. Die rationale Grundlage für diese Entscheidungen muss das IQWiG schaffen, dessen Aufgabe einer wissenschaftlichen Expertise gesetzlich ausdrücklich von der Entscheidung etwa über die Kostenübernahme getrennt ist.

3 EbM: Was ist ein Argument?

Bestimmte Orte und bestimmte Aufgaben lassen bestimmte argumentative Strategien und Inhalte zu, während andere nicht berücksichtigt werden bzw. ausgeschlossen sind. Beim G-BA und bei nationalen Leitlinien ist die EbM explizit als Kriterium für Empfehlungen

4 Der Begriff *Claim* ist international gebräuchlich für „Anspruch in Bezug auf Wirksamkeit, Sicherheit, Angemessenheit, Kosteneffektivität usw. einer Technologie“.

5 Für die gesetzliche Krankenversicherung (GKV) in Deutschland ist tatsächlich die Kostenübernahme nicht entscheidend, da alle zugelassenen Medikamente verordnungsfähig sind. Bei einem vom G-BA festgestellten Zusatznutzen kann ein Anbieter jedoch einen höheren Preis als für die Vergleichstherapie verlangen. Wird dieser Zusatznutzen nicht festgestellt, stellt das pharmazeutische Unternehmen wegen reduzierter Umsatzerwartungen häufig die Vermarktung in Deutschland ein.

bzw. Entscheidungen gefordert. Während David Sackett für die klinische Entscheidung betont, dass neben dem Bezug auf einschlägige und valide Studien die Präferenzen der Patientin und die Erfahrung der Ärztin zu berücksichtigen sind (Sackett et al. 1996), dominiert bei Repertoire-Entscheidungen der Bezug auf Erkenntnisse, die an Populationen gewonnen worden sind. Zwar ist die Bedeutung des englischen Wortes *evidence* sehr breit; sie umfasst alles, was sich als Beleg für oder gegen eine Hypothese, Schlussfolgerung oder Einsicht vorbringen lässt. In unserem Kontext ist seine Bedeutung weitgehend auf den Wissenschaftsbezug eingegrenzt. Im Folgenden diskutiere ich die nach EbM-Maßstäben möglichen und akzeptablen Argumente.

3.1 Fragestellungen und zulässige Studiendesigns, Hierarchie der Evidenz

Zentrales Gliederungselement der EbM-Logik ist die Fragestellung: zunächst muss geklärt werden, ob es um die Wirksamkeit und Sicherheit einer Therapie geht; die Fähigkeit eines diagnostischen Tests, an einer bestimmten Krankheit Erkrankte zu erkennen; die Prognose einer Erkrankung zu erfassen usw. Davon ausgehend kann mit dem methodischen Instrumentarium der EbM geklärt werden, ob die Publikation (oder Studienplanung) geeignet ist, die Frage zu beantworten und damit den Einsatz einer Maßnahme zu rechtfertigen. Für die Wirksamkeit einer Therapie sind randomisierte kontrollierte Studien (RCT) in der Regel eine obligate Voraussetzung, bei einem diagnostischen Verfahren die Querschnittsstudie mit einem Referenzstandard, bei der Prognose eine über eine ausreichende Zeit verfolgte Kohorte von betroffenen Patientinnen usw. Um die einschlägige Literatur zu einer Fragestellung möglichst vollständig zu erfassen, sind Verfahren der systematischen Literaturübersicht entwickelt worden. Diese umfassen die Recherche und den Ein-/Ausschluss einschlägiger Publikationen, die standardisierte Analyse ihrer Charakteristika und – wenn möglich – die Berechnung von Schätzwerten für Wirksamkeit, Sicherheit, diagnostische Treffsicherheit usw., die sog. Metaanalyse. Für die verschiedenen Fragestellungen sind Hierarchien der Belastbarkeit von Evidenz formuliert worden (Howick et al. 2009).

3.2 Patientenrelevante Outcomes

Die Effekte von medizinischen Maßnahmen lassen sich zum einen durch apparative oder laborchemische Messungen erfassen, wie z. B. das Absinken des Blutdrucks, eine Verminderung von Herzrhythmusstörungen im EKG oder eine verminderte Harnsäurekonzentration im Blut. Zum anderen gibt es Kriterien, welche von der Patientin bemerkt werden und direkte Implikationen für Lebensqualität bzw. -dauer haben: das Auftreten eines Schlaganfalls oder Herzinfarkts, eine Episode von Gicht oder eine Verschlechterung des Asthmas, welche Steroide (Kortison) oder gar eine Krankenhaus-Aufnahme erfordert.

Im Kontext der EbM zählen nur Effekte, die in Bezug auf patientenrelevante Outcomes demonstriert worden sind. Innovationen, deren Effekte sich nur in Laborwerten oder technischer Bildgebung („Surrogatgrößen“) äußern, mögen vielversprechend sein. Ihr

Einsatz ist jedoch zu vermeiden, solange sie nicht in Studien mit patientenrelevanten Zielkriterien geprüft worden sind.

Solche Surrogatgrößen sind zwar dem pathologischen Mechanismus „näher“, welchen z. B. ein Medikament beeinflussen soll. Dessen Wirkungen lassen sich auf diese Weise schneller (geringere Studienlaufzeit) und einfacher (kleinere Stichprobe) erfassen – ganz abgesehen von der suggestiven Kraft von solchen Untersuchungen. Mit dem postulierten Mechanismus lässt sich die Wirkung einer medizinischen Intervention auf den menschlichen Organismus jedoch nicht vollständig erfassen. Schädigende Wirkungen an anderer Stelle mögen den erwünschten Effekt neutralisieren oder gar übertreffen. Für das Medikament *Flecainid* war nachgewiesen, dass es im EKG gezeigte Herzrhythmusstörungen verminderte; trotzdem erhöhte es die Sterblichkeit in einer großen Placebo-kontrollierten Studie (Echt et al. 1991). Seitdem gelten Studien auf der Basis von Surrogaten als unzureichend.

3.3 Bias

In methodischen Werken und in Auseinandersetzungen über konkrete Studien ist der Hinweis auf „Bias“⁶ allgegenwärtig. Mehrere Dutzend Arten des Bias sind inzwischen beschrieben (Sackett 1979); grob einteilen lassen sie sich nach Selektions- und Beobachtungs-Bias. Zu ersterem gehören die verzerrte Auswahl einer Stichprobe, Unterschiede zwischen einer Interventions- und einer Kontrollgruppe, Verluste von Patientinnen während der Studienlaufzeit usw. Ein Beispiel für Beobachtungs-Bias wäre etwa eine durch Erwartungshaltung geschönte Angabe von Symptomen oder verfälschtes Ablesen von Messdaten. Das Instrumentarium klinischer Studien kennt vielfältige Vorkehrungen gegen Bias; die wohl wichtigsten kommen im Begriff der „randomisierten, kontrollierten Doppelblind-Studie“ (RCT) zum Ausdruck. Dabei wird eine Kontrollgruppe von unbehandelten Patientinnen mitgeführt, die zum Vergleich mit den Behandelten dient. Die Zuteilung zu den beiden Gruppen wird durch ein verdecktes Zufallsverfahren geleistet, um (auch unbewusste) Manipulationen zu vermeiden. Verblindet sollen sowohl die teilnehmenden Patientinnen als auch das sie betreuende Personal sein, um Erwartungs- und Verfälschungs-Effekte durch das Wissen um die Behandlung („Placeboeffekte“) zu minimieren; heute fordert man auch eine Verblindung des Studienpersonals und der auswertenden Statistikerinnen. Eine Studie kann umso eher als Begründung für eine Maßnahme gelten, je mehr es gelungen ist, in Design, Durchführung und Auswertung möglichen Bias zu reduzieren.

Interessant ist nun, dass die Verfechter der EbM das positive Gegenstück zum Negativ-Kriterium des Bias, nämlich ihren Wahrheitsbegriff, kaum thematisiert haben. Im oben erwähnten ersten Artikel (Department of Clinical Epidemiology and Biostatistics, McMaster University 1981) wird bereits darauf hingewiesen, dass es leichter ist, eine Publikation wegen mangelhafter Validität zu verwerfen als positive Aussagen zur Validität zu machen.

⁶ Im Deutschen werden die Ausdrücke Verzerrung, Verfälschung, Trugschluss usw. kaum genutzt, stattdessen der aus dem Griechischen stammende, englisch ausgesprochene Term „Bias“.

Auch wenn sich andere (Djulbegovic et al. 2009) darum bemüht haben, vermute ich, dass die Angehörigen der McMaster-Gruppe als pragmatisch orientierte Kliniker die Definition eines positiven Wahrheitsbegriffs als wenig aussichtsreiches Unterfangen einschätzten.

3.4 Signifikanz vs. Relevanz

Neben möglichem Bias ist bei einem Studienergebnis die Rolle des Zufalls zu evaluieren. Fast immer wird klinische Forschung an Stichproben durchgeführt. Selbst wenn kein Selektions-Bias gegeben ist, wird sich deshalb bei sämtlichen Schätzungen eine zufällige Abweichung von der Grundgesamtheit ergeben, der sog. Stichprobenfehler. Konkret stellt sich bei einer Studie, bei welcher der Studienarm etwas besser als die mit Placebo behandelten Patientinnen abschneidet, die Frage: ist dies eine zufällige Schwankung, oder sollte man von einer tatsächlichen Wirkung ausgehen? Für diese Frage gibt ein statistischer Test bzw. ein Konfidenzintervall eine Orientierung.

Selbst wenn der Test ergibt, dass es sich bei der Überlegenheit der neuen Behandlung mit hoher Wahrscheinlichkeit nicht um einen Zufallseffekt handelt, ist die Beurteilung damit nicht beendet. Selbst wenn das Ergebnis „statistisch signifikant“ ist, d. h. mit hoher Wahrscheinlichkeit eine Wirkung angenommen wird, muss diskutiert werden, ob diese klinisch relevant ist. Eine sehr große Stichprobe kann nämlich auch sehr geringe Effekte statistisch absichern, die für den Alltag der Versorgung den personellen Aufwand, die Kosten und die möglichen Nebenwirkungen nicht aufwiegen.

Dies erfordert Urteilsvermögen, letztlich geht es um Wertvorstellungen: ab welcher Zahl von verhinderten Schlaganfällen oder Todesfällen sprechen wir von einem „relevanten“ Effekt? Als Nebenwirkung tritt eine seltene, aber schwere und lebensbedrohliche Hautreaktion auf; wie viele (verhütete) Schlaganfälle können diesen Nachteil aufwiegen? Erschwert wird die Beurteilung durch unterschiedliche Perspektiven: Eine Blutdrucksenkung um 2 mmHg ist für die individuelle Patientin nicht relevant; bezogen auf die Bevölkerung eines Landes kann sie einige hundert Schlaganfälle verhüten.

Die Relevanz von Studienergebnissen ist tatsächlich der Punkt, an dem sich aktuelle Kontroversen bevorzugt entzünden. Durch die gültigen Frameworks ist der methodische Spielraum bei der Planung und Durchführung klinischer Studien in hohem Maße eingengt worden, zumindest dann, wenn die Marktzulassung eines Medikaments angestrebt wird. Während in der Frühzeit der EbM eine kritische Frage lautete: Gibt es eine randomisierte-kontrollierte Studie zur Wirksamkeit von X?, stellt dies heute kein relevantes Unterscheidungskriterium dar, da dieses Design heute allgemein akzeptiert ist (Donner-Banzhoff/Lelgemann 2003).

3.5 Externe und interne Validität

Die Wahl des Studientyps, der Durchführung der Maßnahme, Verblindung und Vollständigkeit des *Follow-ups* stellen Aspekte der sog. ‚internen Validität‘ dar. Es geht hier darum, ob die am Ende gezogenen Schlussfolgerungen für die in der Studie untersuchten Personen gültig sind. Mit der ‚externen Validität‘ meinen wir die Verallgemeinerbarkeit

der Befunde auf die Versorgung von Patientinnen bzw. Gruppen jenseits der Studienteilnehmerinnen. Lassen sich die Erkenntnisse aus einer in Kanada durchgeführten Studie auf europäische Patientinnen anwenden? Wird das, was in mehreren Spezialambulanzen funktioniert, auch in der hausärztlichen Versorgung Erfolg haben? Sind Ergebnisse aus dem Studium von Patientinnen mittleren Alters auch auf alte Menschen anwendbar? Diese Extrapolationen erfordern regelmäßig ein hohes Maß an „judgement“, sowohl bei der klinischen Anwendung wie auch der Repertoire-Entscheidung.

3.6 Gesundheitsökonomische Evaluation

Auch wenn die Einschätzung der Relevanz eines statistisch signifikanten Studienergebnisses immer subjektiv bleibt, lässt sich dieser Prozess doch formalisieren. Gesundheitsökonomische Evaluationen fragen, ob eine medizinische Intervention angeboten werden sollte, wenn man die notwendigen Ressourcen bedenkt. Oder fährt eine Gesellschaft besser, wenn sie diese Mittel in andere Maßnahmen oder Programme investiert (Drummond et al. 2007)? Voraussetzung für solche Überlegungen sind Erkenntnisse über die klinische Wirksamkeit, die aus den o. g. Studien gewonnen werden müssen. Dies hat zu dem Vorwurf geführt, dass die EbM zur Kürzung von medizinischen Leistungen (Rationierung) eingesetzt würde.

Gesundheitsökonomische Erwägungen führen zu harten Entscheidungen, etwa ob sehr hohe Mittel zur Rettung *eines* Menschenlebens oder für kleine Zugewinne an Gesundheit einer *großen Zahl* von Menschen aufgewendet werden sollen. Während für die Darstellung der Alternativen ein ausgefeiltes wissenschaftliches Instrumentarium zur Verfügung steht, scheuen politische Entscheidungsgremien verständlicherweise diese Konfrontation. Recht explizit ist die Regel des britischen Gesundheitsdienstes, dass eine durch eine neue Behandlungsmethode erreichte Lebensverlängerung um ein Jahr (bereinigt für die Lebensqualität) nicht mehr als 30 000 £ kosten darf (McCabe et al. 2008). Nur wenige Länder oder Systeme haben Prozeduren für eine Prioritätensetzung nach gesundheitsökonomischen Gesichtspunkten. Stattdessen greifen viele zu verdeckten Formen der Rationierung, sei dies durch aufwändige Genehmigungsverfahren oder die Beschränkung auf eng definierte Gruppen.

3.7 Variierende „Tiefe“

Das Auffahren dieser methodischen Argumente für oder gegen einen *Claim* ist in dieser Tiefe nicht allzu oft finden. Das *Arzneitelegramm* als kritische, nicht von Anzeigen abhängige Zeitschrift vom Typ des „Drug Bulletin“ liefert regelmäßig eine in dieser Weise detaillierte Kritik aktuell publizierter Medikamentenstudien. Auch das IQWiG geht in seinen Berichten sehr gründlich auf die methodischen Charakteristika der einschlägigen Studien ein. Das oben erwähnte Team der Krankenhausärztinnen wird vermutlich auf die aktuelle Leitlinie ihrer Fachgesellschaft, vielleicht noch auf einen Überblicksartikel Bezug nehmen, der zufällig zeitnah erschienen ist (und vielleicht sogar der Anlass für die

Initiative gewesen ist). Beim kollegialen Gespräch in einer Veranstaltungspause ist der Evidenzbezug schließlich auf die Floskel „Neue Studien zeigen...“ geschrumpft.

4 Prozedurale Formalisierungen

Die Abläufe in einem Gremium wie dem G-BA sind durch Gesetz (SGB V) und Geschäftsordnung hochgradig formalisiert. Dabei bildet die EbM den fachlichen Hintergrund, d. h. Publikationen liegen vor, welche daraufhin untersucht werden, ob sie einen *Claim* unterstützen oder nicht. Sowohl beim G-BA und seinen Unterausschüssen wie auch bei den Berichten des IGWiQ sind umfangreiche Anhörungsmöglichkeiten vorgesehen.

Ähnliches gilt für die Leitlinien. Die medizinischen Fachgesellschaften haben de facto ein Monopol für Leitlinien in Deutschland. Das Verfahren wird durch die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) entwickelt, abgestimmt und unterstützt. Auch hier ist der Bezug auf die EbM essenziell: „[Leitlinien] sollten auf einer systematischen Sichtung und Bewertung der Evidenz und einer Abwägung von Nutzen und Schaden alternativer Vorgehensweisen basieren“ (AWMF 2021).

Diese Verfahren bewegen sich in einem Spannungsfeld. Auf der einen Seite beinhaltet die Orientierung an der Evidenz in Form von einschlägigen und validen Studien einen objektiven Anspruch: idealerweise soll aus wissenschaftlichen Aussagen „objektiv“ hergeleitet werden, wie eine Patientin zu behandeln sei. Auf der anderen Seite ist allen mit der Materie Vertrauten bekannt, dass die Komplexität von Gesundheitsversorgungs-Settings – sei dies auf der Mikro- oder der Makro-Ebene – diesen simplen Anspruch konterkariert. Hinzu kommen die unterschiedlichen Perspektiven und Interessen, die immer berührt sind. Während der erste Gesichtspunkt einer wissenschaftlich orientierten Logik entspricht, steht der zweite für den Ausgleich von Interessen, wo dann durchaus Rhetorik ihren Platz beanspruchen kann.

Die meisten der o. g. Verfahren werden direkt oder indirekt von Unternehmen der Pharma- oder Medizinprodukt-Branche in Gang gesetzt. Dies ist staatlich gewollt und wird auf vielen Ebenen gefördert (Ausstattung von Universitäten und wissenschaftlichen Instituten, Patentschutz, Förderung von Start-ups usw.). Auf der anderen Seite erfüllen Verfahren, die im Labor bzw. als Tierversuch vielversprechend erscheinen, in der realen Versorgung nicht notwendigerweise die in sie gesetzten Erwartungen; eine kritische Prüfung mit klinischen Studien ist deshalb unverzichtbar. Seit dem frühen 20. Jahrhundert ist der Schutz der Bevölkerung vor wirkungslosen oder gar schädlichen Behandlungen eine allgemein akzeptierte staatliche Aufgabe geworden. Hinzu kommt, dass selbst die Gesundheitssysteme reicher Länder wegen begrenzter Budgets Prioritäten setzen müssen. Letztlich findet nur ein Bruchteil medizinischer Innovationen den Weg in die Versorgungsroutine.

5 Traditionelle und geächtete Argumentation

Vor dem Hintergrund der durch die EbM veränderten Argumentationsweisen ist die Abgrenzung von deren Alternativen interessant. Diese sind sowohl sequenziell (was ging der EbM voraus?) wie auch parallel (welche alternativen Argumentationsweisen sind heute außerdem gültig?) zu diskutieren.

Halb ironisch, halb ernst ist die EbM zu ihrem Beginn von der sog. Eminenz-basierten Medizin abgegrenzt worden (Isaacs/Fitzgerald 1999). Während eine klinische Studie eine objektive, transparente, von einer einzelnen Person unabhängige Einsicht darstellen soll, weist dieser Ausdruck auf eine Person als Zeugin für die Wirksamkeit bzw. Sicherheit einer Maßnahme hin. Dieses Autoritätsargument wird auch deutlich in Formulierungen wie „Die Kölner machen es soundso, die Kieler dagegen...“.

Eng damit zusammen hängt die Begründung mit der persönlichen Erfahrung. Diese wird durch den Bezug auf die systematisch an einem Kollektiv erhobene Studienevidenz (deklaratives Wissen) relativiert, wenn nicht gar entwertet. Aus der Sicht der EbM sind daran – übertragen in die Begrifflichkeit klinischer Studien – zu kleine Stichproben, un-systematische und verzerrte Datenerhebung sowie unvollständiges *Follow-up* (nicht alle Patientinnen stellen sich nach der Behandlung wieder vor) zu kritisieren. Hinzu kommen Regressionseffekte (spontane Besserung des Krankheitsverlaufs – „physician’s friend“) und selektive Inanspruchnahme der Patientinnen. Andererseits bleibt „Erfahrung“ (prozedurales Wissen) als persönliches Charakteristikum und als Gegenstück von theoretischem Lernen von großer Bedeutung. Dies trifft für Ärztinnen zu, denen klar ist, dass das Studium von Texten nicht ausreicht, um Diagnosen zu stellen oder manuelle Eingriffe verlässlich durchzuführen. Aber auch Patientinnen vertrauen sich lieber Ärztinnen an, die eine Aura von Erfahrung verstrahlen (wie gut dies auch immer begründet sein mag).

Mechanistische Plausibilität stellt eine weitere wichtige Argumentationsweise dar. Ärztinnen werden während ihrer Ausbildung stark in Richtung eines Denkens in (physiologischen, biochemischen) Mechanismen geprägt (Donner-Banzhoff 2022, Kap 15). Ein neues Medikament wird umso eher akzeptiert und in die Routine übernommen, je besser der für seine Wirkung postulierte Mechanismus zu den bereits gehegten Auffassungen zur Funktion des Organs und der Entstehung einer Erkrankung „passt“. Die Studien, welche diese Behauptung begründen, bleiben hier meist im Dunklen; es handelt sich in der Regel um Tierversuche an kleinen Stichproben mit einer im Vergleich zu klinischen Studien wenig konsentierten Methodik (Kimmelman et al. 2014). Interessanterweise taucht eine detaillierte quantitative Argumentation praktisch nie in der Werbung für Medikamente oder Medizinprodukte auf, wohl aber der – oft eindrucklich illustrierte – Verweis auf den Wirkmechanismus.

Auch die emotionale Rührung durch einen Einzelfall ist nicht gestattet, um etwa eine Empfehlung oder Kostenerstattung zu erreichen. Sehr häufig findet man allerdings eine epidemiologische Begründung für ein neues Verfahren, d. h. quantitative Angaben zur Häufigkeit, Dauer und Schwere einer Erkrankung. Diese werden typischerweise auch in

der Einleitung einer wissenschaftlichen Publikation zur Wirksamkeit genannt und stehen in dem Ruf, in der Regel übertrieben zu sein.

Diese Begründungsweisen sind durch die EbM zwar relativiert, bei den o. g. Repertoire-Entscheidungen sogar verpönt; sie werden jedoch weiterhin gepflegt. So sind beide noch in den Werbematerialien der pharmazeutischen und Medizinprodukte-Industrie anzutreffen. Hier werden auch weiterhin „Meinungsführer“ eingespannt, deren Äußerungen von Firmenpublikationen und anzeigenfinanzierten Zeitschriften (Becker et al. 2011) verbreitet werden. Die Plausibilität einer Innovation wird durch eindrucksvolle Visualisierungen des postulierten Mechanismus unterstützt. Völlig andere Gesetze gelten dann, wenn die Diskussion das juristisch und methodisch eingehegte Fachgremium verlässt und in die allgemeine Öffentlichkeit gelangt. Hier wird der bewegende Einzelfall zum wesentlichen, vielleicht sogar einzigen Argument. Dies geschieht in weitgehend staatlichen Gesundheitssystemen (z. B. im United Kingdom) häufig, da hier sämtliche Entscheidungen leicht politisierbar sind – schließlich kann der Gesundheitsminister direkt verantwortlich gemacht werden. Die Selbstverwaltung in kontinentaleuropäischen Systemen (z. B. in Deutschland) dagegen wird von der Öffentlichkeit kaum verstanden; entsprechend gering ist die Anteilnahme an Entscheidungen. Übrigens sind diese Mobilisierungen der allgemeinen Öffentlichkeit keineswegs unschuldig; sie werden dafür verantwortlich gemacht, dass ein hoher Anteil der Mittel für die Behandlung von Krebs und wenigen seltenen Erkrankungen ausgegeben werden, während weniger spektakuläre, eher soziale Randgruppen betreffende Behandlungen, z. B. die Tuberkulose, zu kurz kommen (O'Mahony 2020).

6 Rückwirkungen

Während die EbM vor allem die Rezeption wissenschaftlicher Publikationen betrifft, hat ihre weltweite Etablierung vielfältige Rückwirkungen gehabt.

6.1 Patientinnen-Information und Entscheidungshilfen

Entscheidungshilfen zielen darauf ab, Patientinnen über das anstehende Problem und die verfügbaren Behandlungsoptionen ausgewogen und verständlich zu informieren. Sie sollen den Betroffenen helfen, sich ihrer Präferenzen und Wertvorstellungen klar zu werden und sich auf dieser Grundlage an Entscheidungen zu beteiligen. Digitale Medien bieten dazu vielfältige Möglichkeiten, einschließlich der Integration in die Konsultation mit Gesundheitspersonal. Da die für diesen Zweck nötige Information häufig quantitativ ist (in welchem Maße profitiere ich von einer bestimmten Behandlung? Welche Nebenwirkungen gibt es, und wie häufig sind sie zu erwarten?), werden dafür als Grundlage dieselben klinischen Studien benötigt, welche auch Ärztinnen eine an EbM orientierte Entscheidung ermöglichen (Elwyn et al. 2006). „Shared Decision-Making“ (SDM) entspricht der Entwicklung zu selbstbewussteren, neugierigen und aktiven Patientinnen, auch wenn die Umsetzung im Alltag der Gesundheitsversorgung noch in den Anfängen

steckt. Ohne die EbM bzw. die dafürstehende wissenschaftliche Grundlage ist SDM nicht vorstellbar.

Zunehmend werden systematische Übersichtsarbeiten und klinische Studien von allgemeinverständlichen Zusammenfassungen begleitet, dasselbe gilt für Leitlinien und die internationale Initiative für systematische Übersichtsarbeiten, die Cochrane-Collaboration.

6.2 Produktion von Evidenz

Wenn sich Argumente ändern, welche die „Konsumenten“ wissenschaftlicher Evidenz akzeptieren, hat dies Rückwirkungen auf die „Produzenten“. Hersteller von Medikamenten und Medizinprodukten stellen sich darauf ein, Studien durchzuführen, welche die o. g. Kriterien zumindest näherungsweise erfüllen. Im Ergebnis wissen wir inzwischen wesentlich besser, was medizinische Interventionen leisten und welche Risiken sie mit sich bringen als vor 40 Jahren.

Wenn eine Wirksamkeit auch auf der Ebene patientenrelevanter Outcomes (s. o.) belegt werden muss, sind meist Studien mit mehreren Tausend Patientinnen erforderlich. Je nach Fragestellung müssen diese über mehrere Jahre regelmäßig einbestellt, befragt und untersucht werden. Die Kosten für diese Studien sind enorm und sind einer der Gründe für die Preissteigerungen bei Medikamenten in den vergangenen Jahrzehnten. Zudem sind die formalen und prozeduralen Anforderungen an zulassungsrelevante Studien so hoch, dass diese zu einer Oligopolisierung in der Pharmabranche geführt hat. Die Universität Oxford und das deutsche Unternehmen BioNTech wären nicht in der Lage gewesen, zur klinischen Prüfung ihrer COVID-19-Impfstoffe ein Studienprogramm mit Zehntausenden von Probandinnen durchzuführen. Sie sind deshalb Kooperationen mit etablierten Pharmaherstellern eingegangen, die über das Personal, Know-How und Studiennetzwerke verfügten.

Nur sehr finanzkräftige Akteure sind in der Lage, den Kriterien der evidenzbasierten Medizin entsprechende Evidenz zu generieren. Ein neues Gesprächsverfahren für psychisch Kranke, eine medizinische Entscheidungshilfe oder eine neue Operationstechnik versprechen jedoch keine Umsätze, oft sogar das Gegenteil davon. Wenn kommerzielle Interessen und die entsprechende Investitions-Bereitschaft fehlen, müssen öffentliche Förderer oder private Stiftungen einspringen. Diese Finanzierungsquellen sind wegen geringer Förderquoten jedoch unsicher bzw. in ihrem Volumen begrenzt. Damit entsteht in den Wissensbeständen mit medizinischer Relevanz ein Ungleichgewicht zugunsten pharmakologisch bzw. technisch definierter Technologien mit Vermarktungspotenzial.

Die aktuelle Rechtslage ist paradox: es ist Aufgabe der pharmazeutischen Firma, d. h. des Spielers mit dem stärksten Interessenskonflikt, ihr neues Produkt in Bezug auf Wirksamkeit und Sicherheit zu evaluieren. Diese Paradoxie hat ein detailliertes Regelwerk (*International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use – Good Clinical Practice*) erzwungen, um die ethischen, rechtlichen und wissenschaftlichen Standards an klinische Prüfungen sicherzustellen. Dieses Regelwerk ist

in den meisten Ländern und Regionen verpflichtend für die Marktzulassung von Medikamenten. Es hat zu einer extremen Regelungsdichte und zu Bürokratien geführt, welche dafür sorgen, dass diese Regelungen eingehalten werden. Der Schutz der Gesundheit und der Rechte von Studienteilnehmerinnen sind hohe Güter, aber eine Konzentration in der pharmazeutischen und zunehmend auch Medizinprodukte-Industrie sowie hohe Kosten neuer Behandlungen sind die wohl nicht absehbaren Folgen gewesen.

6.3 Machtverhältnisse

Die Idee und erste Untersuchungen für ein neues Medikament oder Medizinprodukt mögen noch an einer Universität angesiedelt sein. Für das weitere Schicksal einer Innovation ist ein kommerzieller Akteur (Industrie) praktisch unverzichtbar (Latour 1987). Die Etablierung der EbM hat die Unternehmen der Pharma- und Medizinproduktbranche eher gestärkt; denn nur sie können ein Studienprogramm anstoßen und finanzieren, welches den staatlichen Regelwerken und der EbM genügt. Kleine Start-ups, die auf Grund einer biomedizinischen Idee gegründet worden sind, spekulieren geradezu auf die Übernahme durch einen großen Konzern, welcher ihr Produkt zur Marktreife bringt.

Auch die Meinungsbildung innerhalb der Profession hat sich geändert: der Einfluss der Ordinarien ist durch die o. g. Entwicklungen relativiert worden, sie müssen sich der aktuellen Leitlinie beugen. Studierende können durch Recherche in wenigen Minuten feststellen, ob die Lehrinhalte einer Dozentin auf dem neuesten Stand sind.

Klinische Studien nach den aktuellen Standards haben die empirischen Belege von Wirksamkeit und Sicherheit von Interventionen verbessert. Allerdings bleibt der Einfluss der Industrie auf den Bestand und die Verbreitung medizinischen Wissens beträchtlich. Vor allem die Verflechtung mit Fachgesellschaften sorgt dafür, dass die Industrie bei Leitlinienverfahren indirekt „mit am Tisch sitzt“. Über innovative und teure Technologien zu verfügen, gibt einer Fachgruppe Reputation und Macht (Burri 2008). Abhängigkeit der Fachmedien von Anzeigen, Sponsoring von Kongressen, langfristige Beeinflussung durch Vortrags- und Beiratshonorare führen dazu, dass die Vertreterinnen von Fachgesellschaften häufig massive Interessenskonflikte aufweisen. Gleichzeitig sorgen diese Mechanismen dafür, dass in Fortbildungsmedien eine tendenziöse Berichterstattung z. B. über Innovationen vorherrscht (Becker et al. 2011).

7 Kontroversen

7.1 Streitpunkt randomisierte-kontrollierte Studie (RCT)

Die Literatur, die sich – vor allem zu Beginn – kritisch mit der EbM auseinandergesetzt hat, ist zu umfangreich und (fachlich, regional, kulturell) zu vielfältig, als dass sie hier umfassend dargestellt werden könnte. Ich möchte zunächst das Unbehagen diskutieren, das immer noch in Zusammenhang mit der randomisierten kontrollierten Studie geäußert wird.

Eigentlich ist die randomisierte Zuteilung eine höchst geeignete und doch einfache Methode, mit dem Problem des *Confounding* (= Störfaktoren) umzugehen. Sie stellt sicher, dass sich Interventions- und Kontrollgruppe einer Studie in Bezug auf Merkmale, welche das Outcome beeinflussen können, nur im Rahmen des Stichprobenfehlers unterscheiden. Dazu müssen die Störfaktoren noch nicht einmal bekannt, geschweige denn gemessen oder bei der Auswertung statistisch berücksichtigt werden.

Unter der Bezeichnung von „Real Life Evidence“ o. ä. werden immer wieder Alternativen zur RCT gefordert bzw. angeboten. Dabei handelt es sich um Beobachtungsstudien ohne experimentelle Zuteilung zur Prüf- und Kontrollbehandlung, oft unter Verwendung von Routinedaten. Mit statistischen Verfahren wird eine Adjustierung für erhobene *Confounder* vorgenommen. Während die Verfechter solcher Studien typischerweise die RCT wegen ihrer Künstlichkeit und Distanz vom realen Versorgungsgeschehen kritisieren, behaupten sie hier eine überlegene externe Validität bzw. eine Aussage über die Wirksamkeit von neuen Technologien „unter Alltagsbedingungen“.

Methodisch handelt es sich um einen Rückfall hinter den Stand vor der allgemeinen Akzeptanz der RCT. Auch das ausgefeilteste statistische Adjustierungs-Verfahren kann keine unbekanntes, nicht oder unzureichend (sog. *residual confounding*) gemessenen *Confounder* berücksichtigen – im Gegensatz zur Randomisierung. Dieser Nachteil an interner Validität kann durch einen angeblichen Zugewinn an externer Validität nicht aufgewogen werden (Windeler et al. 2008). Dass die Argumentation der „Wirksamkeit unter Alltagsbedingungen“ immer wieder aufkommt, dürfte zum einen intellektuelle Ursachen haben. Viele Wissenschaftlerinnen sträuben sich dagegen, dem Zufall eine konstruktive Rolle im Forschungsdesign zuzubilligen; dies widerspricht zudem der deterministischen Sozialisation in der theoretischen medizinischen Ausbildung (Donner-Banzhoff 2020). Zum anderen ist es aus der Sicht von Firmen und der mit ihnen assoziierten, oft universitären Meinungsführer attraktiv, eine Technologie *vor* dem Abschluss einer Wirksamkeitsstudie zu verbreiten und erst *danach* die Effektivität und Sicherheit zu evaluieren.

7.2 Barrieren und Akzeptanz

An die EbM sind, nachdem die Idee bekannt geworden ist, große Hoffnungen geknüpft worden: eine Durchforstung des diagnostischen und therapeutischen Wirrwarrs, das Ausmerzen irrationaler Reste und kommerzieller Einflüsse, ein rationales und vollständig transparentes Gesundheitswesen. Auf der skeptischen Seite werden die seelenlose Dominanz statistischer Quantifizierung und eiskalte Rationierung befürchtet, eine Kochbuch-Medizin, die Leistungserbringer wie Patientinnen einschnüren würde.

Kaum einer dieser Aspekte hat sich so erfüllt. Die Klage über die eingeschränkte Therapiefreiheit hat sich zusammen mit denen, die sie geäußert haben, in den Ruhestand verabschiedet; junge Ärztinnen sind froh, für ein alltägliches Versorgungsproblem Leitlinien konsultieren zu können. Dass nur wenige die Methodik klinischer Studien, systematischer Übersichtsarbeiten und evidenzbasierter Leitlinien verstehen, müssen wir wohl akzeptieren. Relevante Studien im Sinne der EbM sind heute häufiger als vor 30 Jahren, auch

wenn die deutsche Universitätsmedizin hier immer noch einen deutlichen Rückstand pflegt. Aber weder kommt die präklinische biomedizinische Forschung zu kurz, noch sind Untersuchungen mit qualitativer Forschung verdrängt worden – ganz im Gegenteil. Unser Wissen über die Wirksamkeit und Sicherheit medizinischer Technologien hat dramatisch zugenommen, die heute verfügbaren Medikamente und Medizinprodukte sind also vertrauenswürdiger als dies vor 40 Jahren der Fall war. Aber Machtgefälle und Ungerechtigkeiten in der Gesundheitsversorgung sind dadurch natürlich nicht verschwunden. Die Beziehung zwischen Patientinnen und Ärztinnen wird sich nie ganz „durchrationalisieren“ lassen. Die archaische Heilfunktion ist vielmehr zu reflektieren und zu nutzen (Donner-Banzhoff 2022, Kap. 16).

Vor einigen Jahren sprachen Kritiker noch vom „Missbrauch“ der EbM, wenn beispielsweise kommerzielle Anbieter sich einer EbM-typischen Argumentation bedienen (Ioannidis 2016). Tatsächlich sind die oben dargestellten Argumentationsmuster in sämtlichen sich auf die medizinische Versorgung beziehenden Diskursen so verbreitet, dass wer gehört werden will, die Argumentation der EbM „gebrauchen“ muss.

8 Relativierungen

8.1 Periodisierung

Wir neigen dazu, im Rückblick historische Abgrenzungen vorzunehmen, die bei näherem Hinsehen sich zugunsten einer allmählichen, geografisch und kulturell heterogenen Entwicklung relativieren. Mit identischem Placebo behandelte Kontrollgruppen wurden bereits vor dem II. Weltkrieg in klinischen Studien mitgeführt (Tröhler 2011), die randomisierte Zuteilung finden wir seit den späten 1940er Jahren (Daniels/Hill 1952); der britische Epidemiologe A. L. Cochrane publizierte seine einflussreichen Überlegungen zu „Effectiveness and Efficiency“ im Jahr 1971 (Cochrane 2004). Bis ins frühe 20. Jahrhundert reichen die Anfänge einer behördlichen Regulierung des Medikamenten-Marktes zurück, welche durch den Contergan-Skandal einen zusätzlichen Schub erhielten. Meine Charakterisierung der EbM als „neuartig“ relativiert sich vor diesem Hintergrund.

8.2 Parallele Entwicklungen

Die Auswirkungen der EbM sind kaum von anderen Entwicklungen zu trennen. Die o. g. Veränderungen im Verhältnis von Professionen einerseits und öffentlichen Körperschaften und Gesellschaft andererseits haben sich auch in Bildungs- und Sozialwesen, Landwirtschaft und Lebensmittel-Hygiene gezeigt. Dasselbe gilt für den der EbM zu Grunde liegenden Anspruch an quantitative Rationalität.

Vor allem die digitalen Medien haben die Informations- und Kommunikations-Landschaft grundlegend verändert. Wissen, welches zuvor nur in Universitäts-Bibliotheken und persönlich im Gedächtnis von Expertinnen gespeichert werden konnte, ist jetzt in Form von Datenbanken, Leitlinien, digitalen Zeitschriften, Informationen für Patientinnen usw. verfügbar. Dadurch ist persönliches Wissen nicht überflüssig geworden, aber doch

relativiert. Um eine Herzschwäche zu diagnostizieren, reicht eine Lektüre einer Leitlinie nicht aus; die Anwendung dieses Wissens in der Praxis ist immer noch „Können“, welches auch nonverbales Wissen und Können umfasst. Aber um zu wissen, welche Medikamente die Prognose der Erkrankung verbessern, muss eine Patientin nicht mehr die Koryphäe in einer weit entfernten Stadt konsultieren. Diese orientiert sich inzwischen an denselben Studien- und Leitlinien, welche für die Ärztinnen am Heimatort gelten.

Die meisten der in den letzten 40 Jahren entwickelten Medikamente haben zum Ziel gehabt, die langfristige Prognose von Menschen mit chronischen Krankheiten zu verbessern. Mit anderen Worten: ihr Effekt lässt sich nicht durch die kurzfristige unmittelbare Erfahrung der Patientin evaluieren. Verhütete Herzinfarkte oder Schlaganfälle, oder ein Hinauszögern des Todes lassen sich nur an großen Stichproben mit mehrjähriger Laufzeit prüfen, dies umso mehr, als die beobachteten Effekte gering sind. Diese Entwicklungen sind ein wichtiger biomedizinischer Grund für die Relativierung ärztlicher Erfahrung und die Notwendigkeit von Langzeitstudien mit patientenrelevanten Outcomes.

9 Schlussfolgerung

Die EbM hat das Sprechen und Entscheiden über medizinische Maßnahmen tiefgreifend verändert. Bemerkenswert sind weniger die (vereinzelt) Kritiken und Beschwerden, sondern die weltweite Akzeptanz dieser Argumentationsweise und Wissensform. Dieses beinhaltet inhaltliche bzw. methodische *Kriterien*, mit denen Wirksamkeit und Unbedenklichkeit sich beurteilen lassen, ein zulässiges *Vokabular*, aber auch eine typische *Rhetorik* („neue Studien zeigen“). Schließlich gehören *Institutionen* dazu, welche die Ergebnisse klinischer Studien aufarbeiten und bewerten; aber auch solche, die auf dieser Grundlage Entscheidungen treffen.

Im Ergebnis wissen wir heute mehr über die Wirksamkeit und Risiken medizinischer Maßnahmen als zuvor. Dieses Wissen ist für die fachliche wie auch allgemeine Öffentlichkeit in bisher nicht gekannter Leichtigkeit zugänglich. Angesichts der Vielzahl von innovativen Technologien sind *Claims* mit der EbM in modernen Gesundheitssystemen überhaupt erst entscheidbar geworden.

Die Akzeptanz der Argumentationsweise muss aber nicht unbedingt in aller Zukunft gegeben sein, sie hängt davon ab, dass die Beteiligten sich den Spielregeln unterwerfen, wobei die staatliche Sanktionierung sicher von großer Bedeutung ist.

Ich danke Günter Ollenschläger für sein hilfreiches Feedback zu einer früheren Version dieser Arbeit.

Literatur

AWMF (2021) AWMF-Regelwerk Leitlinien. Online verfügbar unter <https://www.awmf.org/regelwerk/>, zuletzt aktualisiert am 01.08.2020, [zuletzt aufgerufen am 15.08.2023].

- Becker, Annette/Dörter, Fatma/Eckhardt, Kirsten/Viniol, Annika/Baum, Erika/Kochen, Michael M./Lexchin, Joel/Wegscheider, Karl/Donner-Banzhoff, Norbert (2011) The association between a journal's source of revenue and the drug recommendations made in the articles it publishes. In: *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne* 183(5), 544-548. DOI: 10.1503/cmaj.100951.
- Burri, Regula V. (2008) Doing Distinctions. In: *Social Studies of Science* 38(1), 35-62. DOI: 10.1177/0306312707082021.
- Cochrane, Archibald L. (2004) *Effectiveness and efficiency. Random reflections on health services*. London: The Royal Society of Medicine Press (The Rock Carling Fellowship, 1971).
- Daniels, Marc/Hill, A. Bradford (1952) Chemotherapy of pulmonary tuberculosis in young adults; an analysis of the combined results of three Medical Research Council trials. In: *British medical journal* 1(4769), 1162-1168. DOI: 10.1136/bmj.1.4769.1162.
- Department of Clinical Epidemiology and Biostatistics, McMaster University (1981) How to read clinical journals: I. why to read them and how to start reading them critically. In: *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne* 124(5), 555-558. Online verfügbar unter <https://www.cmaj.ca/content/124/5/555>, [zuletzt aufgerufen am 29.11.2023]
- Djulgovic, Benjamin/Guyatt, Gordon H./Ashcroft, Richard E. (2009) Epistemologic inquiries in evidence-based medicine. In: *Cancer control: journal of the Moffitt Cancer Center* 16(2), 158-168. DOI: 10.1177/107327480901600208.
- Donner-Banzhoff, Norbert (2020) Mechanistische Narrative als Mittel der Dissemination: das Beispiel koronarer Technologien. In: *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 63(5), 521-526. DOI: 10.1007/s00103-020-03130-x.
- Donner-Banzhoff, Norbert (2022) *Die ärztliche Diagnose. Erfahrung - Evidenz- Ritual*. 1. Auflage 2022. Bern: Hogrefe AG.
- Donner-Banzhoff, Norbert/Lelgemann, Monika (2003) Ein neuer Maßstab. Aktuelle Studien verlangen veränderte Beurteilungskriterien. In: *Zeitschrift für ärztliche Fortbildung und Qualitätssicherung* 97 (4-5), 301-306.
- Drummond, Michael F./Sculpher, Mark J./Torrance, George W./O'Brien, Bernie J./Stoddart, Greg L. (2007) *Methods for the economic evaluation of health care programmes*. 3. ed., reprint. Oxford: Oxford Univ. Press (Oxford medical publications). Online verfügbar unter <http://www.loc.gov/catdir/enhancements/fy0640/2005279427-d.html>, [zuletzt aufgerufen am 29.11.2023].
- Ebell, Mark H. (2009) How to find answers to clinical questions. In: *American family physician* 79(4), 293-296.

- Echt, Debra S./Liebson, Philip R./Mitchell, L. Brent/Peters, Robert W./Obias-Manno, Dulce/Barker, Allan H. et al. (1991) Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. In: *The New England journal of medicine* 324(12), 781-788. DOI: 10.1056/NEJM199103213241201.
- Elwyn, Glyn/O'Connor, Annette/Stacey, Dawn/Volk, Robert/Edwards, Adrian/Coulter, Angela et al. (2006) Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. In: *BMJ (Clinical research ed.)* 333(7565), 417. DOI: 10.1136/bmj.38926.629329.AE.
- Freidson, Eliot (1995) *Profession of medicine. A study of the sociology of applied knowledge*. [Nachdr.]. Chicago: Univ. of Chicago Press.
- Guyatt, Gordon (Hrsg., 2002) *Users' guides to the medical literature. A manual for evidence-based clinical practice*. Evidence Based Medicine Working Group. Chicago, Ill.: American Medical Assoc (JAMA & Archives Journals).
- Hannken-Illjes, Kati (2018) *Argumentation. Einführung in die Theorie und Analyse der Argumentation*. 1. Auflage. Tübingen: Narr Francke Attempto Verlag (narr studienbücher). Online verfügbar unter DOI:10.1007/s10503-018-9466-x
- Howick, Jeremy/Glasziou, Paul/Aronson, Jeffrey K. (2009) The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute? In: *Journal of the Royal Society of Medicine* 102(5), 186-194. DOI: 10.1258/jrsm.2009.090020.
- Ioannidis, John P. A. (2016) Evidence-based medicine has been hijacked: a report to David Sackett. In: *Journal of clinical epidemiology* 73, 82-86. DOI: 10.1016/j.jclinepi.2016.02.012.
- Isaacs, David/Fitzgerald, Dominic (1999) Seven alternatives to evidence based medicine. In: *BMJ (Clinical research ed.)* 319(7225), 1618. DOI: 10.1136/bmj.319.7225.1618.
- Kimmelman, Jonathan/Mogil, Jeffrey S./Dirnagl, Ulrich (2014) Distinguishing between exploratory and confirmatory preclinical research will improve translation. In: *PLoS biology* 12(5), e1001863. DOI: 10.1371/journal.pbio.1001863.
- Latour, Bruno (1987) *Science in action. How to follow scientists and engineers through society*. [Nachdr.]. Cambridge, Mass.: Harvard Univ. Press [reprint].
- McCabe, Christopher/Claxton, Karl/Culyer, Anthony J. (2008) The NICE cost-effectiveness threshold: what it is and what that means. In: *Pharmacoeconomics* 26(9), 733-744. DOI: 10.2165/00019053-200826090-00004.
- O'Mahony, Seamus (2020) *Can medicine be cured? The corruption of a profession*. London: head of zeus.

- Oxman, Andrew D./Sackett, David L./Guyatt, Gordon H. (1993) Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. In: *JAMA* 270(17), 2093-2095.
- Sackett, David L. (1979) Bias in analytic research. In: *Journal of chronic diseases* 32(1-2), 51-63. DOI: 10.1016/0021-9681(79)90012-2.
- Sackett, David L./Rosenberg, William M./Gray, J. Muir/Haynes, R. Brian/Richardson, W. Scott. (1996) Evidence based medicine: what it is and what it isn't. In: *BMJ (Clinical research ed.)* 31(7023), 71-72. DOI: 10.1136/bmj.312.7023.71.
- Tröhler, Ulrich (2011) Adolf Bingel's blinded, controlled comparison of different anti-diphtheric sera in 1918. In: *Journal of the Royal Society of Medicine* 104(7), 302-305. DOI: 10.1258/jrsm.2011.10k074.
- Wennberg, John/Girtelsohn, Allenl (1982) Variations in medical care among small areas. In: *Scientific American* 246(4), 120-134. DOI: 10.1038/scientificamerican0482-120.
- Windeler, Jürgen/ Antes, Gerd/Behrens, Johann/Donner-Banzhoff, Norbert/Leigemann, Monika (2008) Randomisierte klinische Studien (RCT). In: *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 102(5), 321-325. DOI: 10.1016/j.zefq.2008.05.002.