

Wie evident ist die Evidenzbasierung ?

Jürgen Kriz, Universität Osnabrück

Zentrale Aspekte eines Vortrags auf der Klausurtagung der WAKS am 13. Februar 2007 in Münster



Das Grundanliegen der „evidenzbasierten Medizin“, EbM, ist fraglos positiv zu bewerten: Es geht nämlich darum, Ärzten (bzw. Psychologischen Psychotherapeuten (PP), Kinder- und Jugendlichen Psychotherapeuten (KJP)) wissenschaftliche Informationen zur Verfügung zu stellen, damit diese integriert mit ihrer ärztlichen Erfahrung und der individuellen Patientensperspektive dem jeweiligen Patienten die beste Behandlung anbieten können.

Kern des dahinter stehenden Anliegens ist es, mit EbM die zunehmenden Schwierigkeiten zu bewältigen, dass „Kliniker“ die Informationsfülle über Behandlungsmethoden (und Medikamente etc.) allein nicht mehr bewältigen können – und zwar sowohl hinsichtlich der Menge an Behandlungsmethoden und Publikationen darüber, als auch hinsichtlich der Gütebeurteilung der meist aufwändigen Studien.

Besonders von der internationalen „Cochrane Collaboration“ werden unter Mitarbeit internationaler Wissenschaftler für diesen Zweck Bewertungen und Zusammenfassungen (sog. „Reviews“) von wissenschaftlichen Studien im Bereich der Medizin vorgenommen.

Ob dieser nützliche Überblick über real durchgeführte Forschungen nach bestimmten Kriterien (s.u.) nun aber auch ein nützliches Kriterium dafür sein kann, welche Methoden in der Praxis angewendet bzw. unterlassen werden sollten, hängt von der Erfüllung bestimmter Voraussetzungen ab. Dazu zählt insbesondere die Frage, wie weit das Spektrum realer Behandlungen tatsächlich und hinreichend biasfrei im Spektrum dieser Forschung widergespiegelt wird. Und hier gibt es aufgrund von Interessenkonflikten erhebliche Zweifel. Denn Forschung richtet sich nicht nur wertfrei an inhaltlich relevanten Fragen der Kliniker bzw. an zu klärenden Fragen für die Patienten aus, sondern:

Für Forschung von Medikamenten, Apparate-Medizin etc., die weitgehend von Firmen finanziert werden, gilt:

- Es werden überwiegend Therapien erforscht, die patentierbar und gewinnversprechend sind;
- begünstigt werden pharmazeutische Unternehmen, die mit wenigen Medikamenten großen Umsatz machen. Diese sind die Gewinner im EbM-Biotop
- Verlierer sind dagegen Therapien ohne Aussicht auf breite Vermarktung oder ohne finanzstarke industrielle Rückendeckung, wie z.B. nicht-pharmakologische Therapien.¹

Für Forschung von Psycho- und Sozialtherapie, Reha-Maßnahmen etc., die primär über universitäre/ öffentliche Mittel (incl. DFG, Stiftungen etc.) finanziert werden, gilt:

- Es werden bevorzugt Fragestellungen erforscht, die im Rahmen von Diplom-, Doktor- und Habilitationsarbeiten angegangen werden und vergleichsweise schnell und einfach publiziert werden können, d.h. die in die universitären Karrierestrukturen passen.
- Beispielsweise lässt sich eine spezielle Vorgehensweise (die dann dem Spektrum „VT“ zugerechnet werden kann) an einer speziellen Patientengruppe sehr gut im Rahmen einer Dissertation experimentell untersuchen. Die Wirkung einer langfristigen, kaum manualisierbaren Vorgehensweise – wie etwa der Langzeit-Psychoanalyse – lässt sich in diesem Rahmen praktisch so nicht untersuchen.

Der durchaus nützliche Überblick über die Forschungsliteratur spiegelt somit keineswegs nur die Nützlichkeit von Interventionsmethoden wider, sondern in erheblichem Ausmaß die Nützlichkeit der Forschung unter kommerziellen und/oder Wissenschaftskarriere-Aspekten.

Zwei wichtige Bestandteile des EbM sind (a) eine bestimmte methodische Vorgehensweise, nämlich die prospektive randomisierte, kontrollierte Doppelblindstudie (randomized controlled trial), RCT, und die ihr zugrunde liegende epistemologische Sicht eines „wissenschaftlichen Beweises“, sowie (b) die systematische Zusammenfassung solcher Studien im Rahmen von Metaanalysen, wobei die Effektstärke (ES) eine besondere Rolle spielt. Diese ist üblicherweise definiert und operationalisiert als Differenz des Mittelwertes der Behandlungsgruppe (Mb) und einer Kontrollgruppe (Mk) hinsichtlich der Post-Messung auf einer Testvariablen, gebrochen durch die Standardabweichung in der Kontrollgruppe (Sk), also:

$$ES = (Mb - Mk) / Sk$$

Diese Kunstgröße ES ist notwendig, damit eben ganz unterschiedliche Testvariablen in ganz unterschiedlichen Gruppen bei ganz unterschiedlichen Fragestellungen irgendwie formal zusammengezogen werden können. Dieser Notbehelf – denn ohne ES wären Metaanalysen aufgrund der Unterschiedlichkeit sehr komplex zu diskutieren – hat sich rückwirkend zunehmend auch als eine Art Gütemaß für einzelne Studien etabliert, weil der Begriff „Effektstärke“ für den Kliniker plausibel und verlockend klingt. Doch sind damit problematische Assoziationen verbunden: Es geht in Wirklichkeit um statistische Diskriminierung der beiden Gruppen. Diese ist groß, wenn Mb-Mk gross sind und/oder Sk klein ist. Während ersteres (unter üblichen Bedingungen) noch tatsächlich grob etwas mit klinischer Effektivität zu tun hat, gilt dies für die letztere nicht: Die klinische Effektivität z.B. einer Suchtbehandlung bei bestimmten Patienten hängt – inhaltlich- schwerlich von der Homogenität der Varianz in einer Kontrollgruppe ab; statistisch aber wird sie doppelt so groß, wenn Sk halb so groß wird (vgl. Kriz 2006). In den Debatten wird ES aber oft so behandelt, als würde diese statistische Diskriminierfähigkeit tatsächlich einen klinisch-inhaltlichen Effekt ausdrücken.

Umstrittener in der Literatur ist die Sinnhaftigkeit von RCT jenseits von Medikamentenstudien, in denen von vornherein eine zufällige Zuteilung zu einer Behandlungsgruppe und einer, die nur ein Placebo erhält, hinreichend gut realisiert werden kann. Die Problematik für andere Bereiche wird klar, wenn man der Zusammenfassung Seligman (1997)² folgt, der zunächst viele Jahre selbst vehement den „Goldstandard“ der RCT „Efficacy“ auch für die Psychotherapie vertreten hatte, inzwischen aber dafür plädiert, zumindest gleichberechtigt auch die *Wirksamkeit in der Realsituation* („Effectiveness“), d.h. die reale Gesamtwirksamkeit der Behandlung, zu berücksichtigen. Er schreibt:

„Grundsätzlich wird dabei eine Behandlungsgruppe unter gut kontrollierten Bedingungen mit einer Kontrollgruppe verglichen. Aber neben der Kontrollgruppe weisen Effizienzstudien, die mittlerweile zu einem paradigmatischen Unterfangen geworden sind, noch eine hochkomplexe Methodologie auf. Im Idealfall gehören die folgenden Raffinessen dazu:

- Zuordnung der Patienten zu der Behandlungs- und Kontrollgruppe nach dem Zufallsprinzip.
- Strenge Kontrollen: Neben Kontrollgruppen ohne Behandlung werden Kontrollgruppen gebildet, bei denen eine Placebo-Behandlung mit potentiell therapeutischen Elementen eingesetzt wird, die für Patienten wie Therapeuten glaubwürdig sind. Damit wird der Einfluss unspezifischer Faktoren überprüft, zum Beispiel Rapport, Erfolgserwartung und teilnahmevolle Aufmerksamkeit (sog. Nonspezifika).

- Die Behandlung ist standardisiert, mit hochgradig detaillierten expliziten Therapievorgaben. Die Einhaltung der Standards wird anhand von Videoaufnahmen der Sitzungen überprüft, Abweichungen werden korrigiert.
- Die Behandlung erfolgt in einer festgelegten Zahl von Sitzungen.
- Die Zielergebnisse werden sorgfältig operationalisiert (z.B. klinisch diagnostizierte DSM-IV-Störungen, Zahl der berichteten Orgasmen, Selbstauskünfte über Panikattacken, prozentueller Anteil flüssiger Äußerungen).
- Auswerter und Diagnostiker wissen nicht, aus welcher Gruppe die Patienten kommen. (Anders als bei Medikamentenstudien sind "doppel-blinde" Effizienzstudien in der Psychotherapie unmöglich, da Patienten und Therapeuten die Behandlungsform kennen. Wann immer Sie jemanden eine "Doppelblindstudie" in der Psychotherapie verlangen hören, halten Sie Ihren Geldbeutel zurück!)
- Die Patienten erfüllen die Kriterien für eine einzige diagnostizierte Störung; Patienten mit mehreren Störungen werden in der Regel nicht einbezogen.
- Nach Behandlungsende werden die Patienten über einen festgelegten Zeitraum mit einer umfassenden Testbatterie eingeschätzt.“ (Seligman 1997).²

Seligman berichtet dann aber weiter, dass ihn die Beteiligung an der „consumer report study“ von 1995 eines besseren belehrt habe. Angesichts der Ergebnisse dieser großen amerikanischen Verbraucher-Umfrage über die Wirkungen von Behandlungen im realen psychotherapeutischen Feld, die sich unbestritten (selbst bei Einräumung aller methodischer Probleme – von denen allerdings auch die Laborforschung, wie alle Forschung, nicht frei ist) sehr deutlich von den aus Laborstudien abgeleiteten Behauptungen über die Wirksamkeit unterscheiden, schwor er seinem bisherigen Irrtum öffentlich ab und machte klar, dass die wissenschaftliche Debatte selbst über empirisch und quantitativ nachgewiesene Effektivität zumindest um diese wichtige Perspektive bereichert werden muss, wenn sie nicht völlig einseitig an der Realität vorbeiarargumentieren will.

So schreibt er: „Durch diese Studie wurde mir klar, dass die Entscheidung, ob eine bestimmte Methode unter streng kontrollierten Bedingungen einer anderen Methode oder der Kontrollgruppe überlegen ist, etwas anderes ist als die Frage nach dem, was in der Praxis wirkt (...). Ich halte „Efficacy“-Studien seitdem nicht mehr für die einzige, ja nicht einmal für die beste Möglichkeit, um festzustellen, welche Methoden in der Praxis tatsächlich wirksam sind. Ich bin zu der Überzeugung gekommen, dass "Effektivenss“-Studien, mit denen die Erfahrungen der Patienten unter den realen Bedingungen in der Praxis erhoben werden, eine brauchbare und glaubwürdige "empirische Validierung" von Psychotherapie und Medikation ermöglichen“ (Seligman 1997²).

Es kann also nicht – wie leider oft verkürzt diskutiert – um die Frage: „RCT: ja oder nein“ gehen. Vielmehr geht es darum, die Bedeutung der RCT richtig einzuschätzen – und zwar im doppelten Sinne von „Bedeutung“: RCT ist bedeutsam für Aussagen, wo das klassische experimentelle Design passt – das gilt im Bereich der Psychotherapie für viele Grundlagenfragen, bei denen spezifische Interventionen experimentell geprüft werden können. Je komplexer allerdings der Bedeutungskontext wird, desto schwieriger wird es, eine angemessene experimentelle Repräsentation dafür zu finden – oder andersherum: desto eher werden fundierte RCT-Aussagen dafür bedeutungslos.

Oft wird allerdings übersehen, dass bereits grundsätzliche Vorannahmen in der Fragestellung selbst stecken. Die – scheinbar neutrale - Frage nämlich: „wirkt Therapiemethode X tatsächlich?“ läuft, bei seriöser und sorgfältiger Beantwortung, unweigerlich auf die o.a. Kriterien der RCT-Studie hinaus: Denn es muss sichergestellt werden, dass dann auch wirklich X geprüft wird, was eine standardisierte, manualisierte und in der Zeit gleiche Intervention ver-

langt, wie in einer Kontrollgruppe. Sofern man nicht annimmt, dass X bei allen gleich wirkt, bedarf es einer möglichst homogen definierten Gruppe, für die X denn wirken soll etc. Die implizite Vorannahme in der Fragestellung ist allerdings, dass es so etwas wie „Therapiemethode X“ als eine Art operationalisierter Input „wirklich gibt“ (oder es zumindest Sinn macht, das zu unterstellen). Genau hier sind aber erhebliche Zweifel für die Psychotherapie angebracht (im Gegensatz zur Pharmaforschung, wo X hergestellt werden kann). Denn „Therapiemethode X“ meint zwar eine hinreichend abgrenzbare Referenz auf ein bestimmtes Menschenbild, Epistemologie, Störungslehre, Veränderungs- bzw. Wirkprinzipien und darauf aufbauend ein spezifisches Spektrum an Interventionsmöglichkeiten. Die konkreten Vorgehensweisen aber sind oft kaum standardisierbar, interaktiv sich entwickelnd und auf recht individuelle Veränderungsprozesse abgestimmt. Die in RCT-Studien „bewiesene“ Wirksamkeit sehr spezifischer Vorgehensweisen an sehr spezifischen Patienten unter sehr spezifischen Bedingungen sind dabei wertvolles und wichtiges Basiswissen und Orientierungshilfen. Aber sie machen eben keine Aussagen über das Abstraktum „Therapiemethode X“, das es real gar nicht gibt (im Vergleich zur realen Existenz einer Pille mit Wirkstoff X).

Ich habe in unterschiedlichen Kontexten (z.B. Kriz 1996³, 2006⁴) Fragen zusammengestellt, die für eine seriöse Beantwortung der o.a. Frage – übereinstimmend! - geklärt sein müssten, von denen allerdings die ganz überwiegende Anzahl ungeklärt bleibt (viele werden dann zwar für die jeweilige Studie in spezifischer Weise geklärt, doch gibt die Studie dann eben Antworten auf diese spezifischen Fragen und nicht über „Therapieform X“). Diese sollen hier zumindest aufgezählt werden – für eine tiefer gehende Diskussion sei auf die o.a. Literatur verwiesen:

Frage: “Wirkt eine bestimmte Psychotherapie überhaupt (und wenn ja: wie effektiv)”?

- wie wird “Therapie-Erfolg” überhaupt definieren?
z.B. “chronischer Krankheiten”
- Wie werden diese Konzepte konkret operationalisiert und gemessen?
(Devereux 1992, S. 27) Sind Fragebogen-Daten eher als Meinungen
oder klinische Diagnosen zu werten?
- An welchen Patienten werden die Messungen überhaupt durchgeführt?
 - Reicht es, eine bestimmte “störungsspezifische” Gruppe” zu wählen?
 - klinischen Klassifikationen von Krankheitsverläufen =
optimale Klassifikationen für Therapieverläufe ?
 - homogene Symptome ?
 - vs. homogene Schwere von Beeinträchtigungen ?
 - vs. homogener Verlauf der bisherigen Krankengeschichten ?
 - vs. homogener “sozialer Support” ?
- Welche Therapeuten ?
 - möglichst repräsentative Auswahl? (Woher Basisinformation?)
 - möglichst erfahrene Therapeuten? (was heißt das?)
wie realistisch, dass erfahrene Therapeuten genau “Methode A” in
“Reinkultur” vertreten?
Wählt man jene, die man gerade bekommen kann?

“Zwischenbilanz”: Antwort nicht: „ob und wie effektiv Therapiemethode A wirkt?“ sondern Information darüber, wie Patienten der Gruppe X, nach den Kriterien C, D und E... zusammengestellt, hinsichtlich Antworten und Beobachtungen der Variablen F, G und H... bei Therapeuten mit den Eigenschaften I, J und K... sich verändern.
Beantwortet das aber noch die obige Frage - bzw. ist diese Information noch als angemessene Basis anzusehen, um Aussagen über “die Therapieform A” zu machen?

weitere Fragen: Vergleich "vorher"/"nachher" (d.h. ein pre-post-design) durchführen, oder aber eine Verlaufsstudie (panelanalyse), bei der zu vielen Zeitpunkten die Information erhoben wird?

Wie soll eine Kontrollgruppe zusammengesetzt sein, die den Therapie-Erfolg gegen Zufallseinflüsse bzw. Spontanremission sichert

Weit schwieriger und unklarer: Vergleich „Therapie A“ mit „Therapie B“

Zusätzlich (!) stellt sich die Frage, hinsichtlich welcher Aspekte (bzw. welches Aspektraumes) ein solcher Vergleich überhaupt stattfinden könnte. Erinnern wir uns: Die obigen Entscheidungen führten sinnvoller Weise zu einer, für A ganz spezifischen Auswahlkombination aller dieser Entscheidungsaspekte. Wenn dieselbe Entscheidungssequenz analog für Therapieform B durchgeführt wird, bleiben dann überhaupt noch irgendwo Überschneidungen, um über einen Vergleich etwa Aussagen zu können?

Meta-Analysen

Welche Studien werden einbezogen?

Wie begründet sich die Auswahl?

Wie wird mit unterschiedlichen Stichproben, mit unterschiedlicher Anzahl von Erfolgskriterien

und den unterschiedlichen gegenseitigen Abhängigkeiten der Kriterien umgegangen?

Generell sind Labor-Studien durch mindestens folgende bedeutsame aber unkontrollierte Aspekte in Bezug auf die klinische Realität verzerrt:

a) selektive Auswahl klinisch valider Patienten (z.B. Ausschluss „komorbider“ oder besonders schwieriger Patienten)

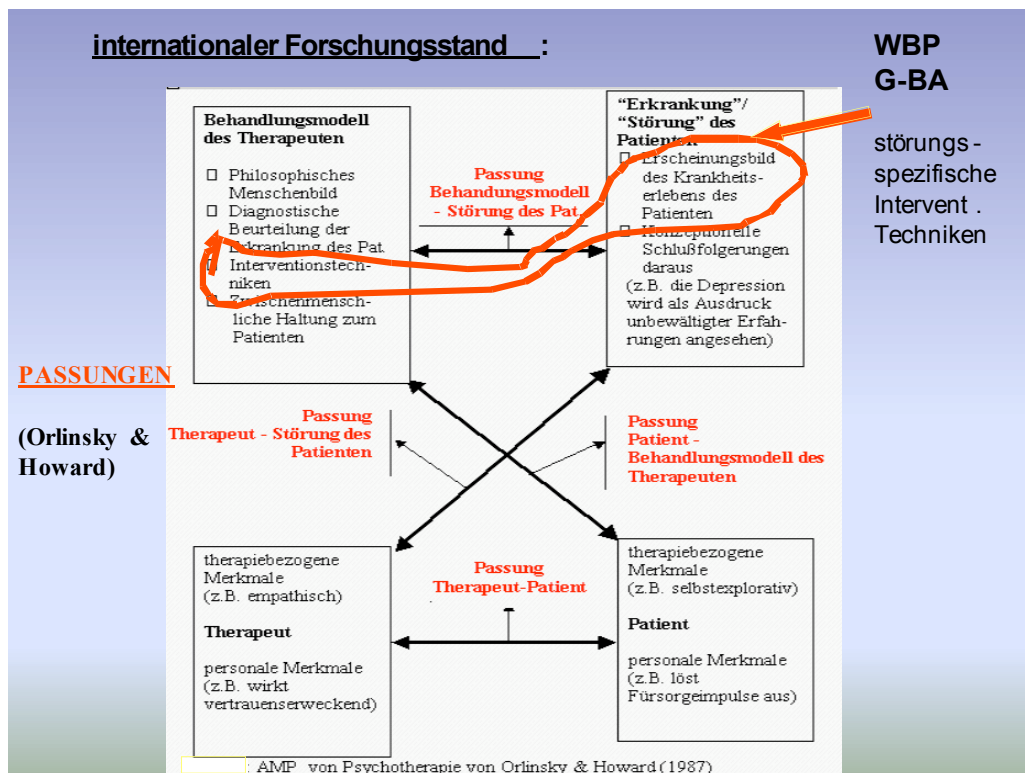
b) selektive Auswahl von Therapeuten (also Ausschluss aller jener, die sich aus Kompetenz oder anderen Gründen nicht an Manuale halten, die mehr /oder weniger/ motiviert sind als die Labor-Therapeuten, etc.)

c) nicht-randomisierte + nicht-repräsentive Auswahl der Therapeuten (d.h. aus all jenen Therapeuten, die Methode A repräsentieren)

d) nicht-Repräsentativität für die real durchgeführten Vorgehensweisen (T_n und andere T_i aus dem Spektrum jener T , die für Methode A (oder gar für VA) aggregiert werden)

Ein weiteres Problem von RCT in der Psychotherapie ist, dass es eigentlich heute zum Minimalstandard einer angemessenen komplexen Diskussion gehört, zumindest das Problem der „Passung“ nach dem Allgemeinen Modell von Psychotherapie (AMP) nach Orlinsky & Howard (1987)⁵ zu beachten. Dieses, angesichts der Komplexität in der Praxis immer noch recht einfache kategoriale Modell, berücksichtigt zumindest folgende Passungen (vgl. Abb)

- Die Passung Behandlungsmodell und Störungsmodell.
- Die Passung Patient und Behandlungsmodell
- Die Passung Therapeut und Patient
- Die Passung Therapeut und Störung des Patienten



• Abb. 1: AMP Allgemeines Modell von Psychotherapie Orlinsky & Howard (1987)

Sieht man sich dazu die faktische RCT- und EbM orientierte Forschung an, so ist diese nicht nur ausschließlich auf den ersten Aspekt ausgerichtet, sondern dies geschieht noch reduziert auf eine bestimmte Methodik, die nicht einmal unter dieser einseitigen Sichtweise notwendig beschränkt sein müsste. So sollte es nachdenklich stimmen, wenn selbst ein Vertreter der EbM, wie Prof. Norbert Schmacke⁶, sich für eine Berücksichtigung qualitativer Methoden einsetzt, die z.B. beim WBP¹ und G-BA² ausgeblendet werden. Ferner wird faktisch eben selbst in den oberen beiden Kästen i.W. nur die beiden umrandeten Aspekte berücksichtigt.

Bemerkt werden soll auch, dass das in manchen Debatten vermittelte Bild, die Psychotherapie müsse sich eben dem Medizin-Sektor anpassen, wo längst EbM allgemeiner Standard sei, so nicht zutrifft:

Der deutsche „Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen (SVR)“ stellte noch in seinem Gutachten von 1999 klar, dass eine "Inventur des Evidenz-Bestands" der Medizin sehr ernüchternd gewirkt habe, "weil sie aufzeigte, dass die Bereiche gesicherten Wissens in der Medizin sehr viel kleiner und die Grauzonen sehr viel größer sind als erwartet". Gemäß einer Übersicht von 1992 zeigte sich, dass "nur etwa 4% aller ambulant und stationär erbrachten Dienstleistungen dem Anspruch auf belastbare Evidenz genügen, 45% genügen einfacheren Evidenzkriterien und für den 'Rest' (rechnerisch 51%) gibt es heute keine wissenschaftliche Evidenz" (SVR 1999, S. 79).⁷

Im Gegensatz zu der im WB vertretenen Richtlinien-Psychotherapie, die ein strikt ätiologisches Krankheitsmodell fordert (Faber/Haarstrick 1999, S. 11 ff.) und eine "konditionalisti-

¹ WBP = Wissenschaftlicher Beirat Psychotherapie (nach & 11 Psychotherapeutengesetz PsychThG)

² G-BA = Gemeinsamer Bundesausschuss (Der Gemeinsame Bundesausschuss ist ein Gremium der gemeinsamen Selbstverwaltung von Ärzten, Krankenkassen und Krankenhäusern. Seine Aufgabe ist es zu konkretisieren, welche ambulanten und stationären medizinischen Leistungen ausreichend, zweckmäßig und wirtschaftlich sind und somit zum Leistungskatalog der Gesetzlichen Krankenversicherung gehören.)

sche Betrachtungsweise" (S. 13) von Krankheit ausdrücklich ablehnt, kommt der Sachverständigenrat zu einer ganz anderen Gewichtung: *"Vor diesem Hintergrund bedarf das ätiologische Krankheitsverständnis einer Revision zugunsten eines konditionalen Krankheitsverständnisses, nach dem Krankheit wesentlich von den Lebensumständen, vom Lebensstil und der patientenseitigen Interpretation abhängt"* (SVR 1999, S. 68).

In diesem Kontext weist der SVR auf die generellen Probleme der Medizin als angewandter Naturwissenschaft hin und fordert explizit eine Ergänzung der evidence-based medicine durch die narrative-based medicine (NbM) (Greenhalgh 1999⁸), die sich nicht auf Naturwissenschaften, sondern auf Phänomenologie und Hermeneutik stützt.

"Die Anwendung (der) Erkenntnisse verleiht ihr jedoch erst in einem System völlig anderer Denk- und Bewertungsstrukturen, nämlich dem Leben des individuellen Patienten und dem gesellschaftlich kulturellen Umfeld, ihren Sinn" (SVR 1999, S. 66).

"In der NbM wird ärztliches Handeln als ein deutender Vorgang betrachtet, der narrative Fähigkeiten erfordert, um die 'Geschichten' der Patienten und der Kliniker mit objektivierbaren medizinischen Befunden (z.B. Testergebnissen) zu verbinden. Die Akkumulation fallbezogener Erfahrungen (case expertise) befähigt den klinisch tätigen Arzt, bei der klinischen Entscheidungsfindung die angemessenste medizinische Maxime auszuwählen. Unter case expertise ist hierbei die Fähigkeit des Arztes zu verstehen, die Geschichten und die 'Krankheitsskripte' (aber auch die klinischen Anekdoten der Kollegen) zu verstehen und zutreffend zu deuten. Schwierigkeiten der Ärzte, wissenschaftliche Erkenntnisse im Kontext der klinischen Begegnung mit den Patienten angemessen umzusetzen, entstehen aus der Sicht der NbM vor allem dann, wenn das narrativ-deutende Paradigma aufgegeben wird und sich klinisches Handeln ausschließlich auf externe wissenschaftliche Evidenz stützt" (SVR 1999, S. 67).

Entsprechend ergänzend zum ätiologischen Modell schlägt der SVR systemtheoretische Ansätze (z.B. das Situationskreismodell von v.Uexküll und Wesiack), insbesondere aber auch lebenswelttheoretische bzw. phänomenologisch-hermeneutische Ansätze vor.

"Diese Denkrichtungen gehen über den systemischen Ansatz hinaus, indem den Prozessen im System nur eine Funktion innerhalb eines jeweiligen unter Umständen sehr engen Kontextes zugewiesen wird. Es geht darum, den subjektiven Krankheitswert und damit die Ebene des Krankheitserlebens des Patienten, ebenso wie die realen Konsequenzen einer Krankheit einzubeziehen. Die Bemühungen der Medizin, Lebensqualität als Folge von Krankheit oder auch einer Therapie zu erfassen, lassen sich hier zuordnen" (SVR 1999, S. 69).

In einem neuerlichen Gutachten kommt denn auch der SVR zu folgender Aussage, was die „evidenzbasierten Leitlinien“ betrifft – Handlungsempfehlungen für Ärzte (und Patienten), um die wissenschaftlichen Erkenntnisse auch tatsächlich umzusetzen:

Es „besteht das prinzipielle Problem, dass die auf populationsbezogenen empirischen Befunden basierenden Leitlinien in ihrer praktischen Anwendung auf den individuellen Patienten bezogen werden müssen. ... Randomisierte, kontrollierte klinische Studien, die lediglich die Effektivität (und gegebenenfalls die Effizienz) einer Maßnahme unter artifiziellen Studienbedingungen („efficacy“) beschreiben, werden in vielen Fällen überbewertet“ SVR, Gutachten 2003, S. 87⁹)

Letztlich sei noch darauf verwiesen, dass die rd. 60 Leitlinien der „Arbeitsgemeinschaft der wissenschaftlichen medizinischen Fachgesellschaften (AWMF)¹⁰ im Bereich Psychiatrie und Psychotherapie schon teilweise Psychotherapierichtungen den *Ärzten* und ihren Patienten empfehlen, die in Deutschland für *Psychologische* Psychotherapeuten noch mit dem Bann der Berufsverbote belegt sind (wie die systemische und Familientherapie) oder dies bis vor kurzem waren (wie der personenzentrierte Ansatz nach Rogers) – ganz abgesehen davon, dass diese Verfahren, die gegenwärtig in der BRD kein *Psychologischer* Psychotherapeut lernen kann

bzw. darf, an (*medizinischen*) *Kliniken* nach wie vor in erstaunlich hohem Maße eingesetzt werden.

So heißt es in der vom AWMF erstellten Leitlinie zur „Paar- und Familientherapie“ (die bisher ebenfalls in Deutschland vom WB nicht als „wissenschaftlich“ anerkannt wurde – und daher für Psychologen als Psychotherapie nicht durchgeführt werden darf und somit z.B. auch kaum weiter erforscht werden kann):

„Die strikte Ausrichtung an Methoden der im engeren Sinne quantifizierend empirisch arbeitenden Forschung, insbesondere die Zugrundelegung randomisierter kontrollierter (experimenteller) Studien wird in jüngster Zeit kritisch gesehen und auch durch Vorstandsmitglieder der American Psychological Association (APA) selbst schon relativiert (vgl. Kriz, 1999, 2000). Offenbar wurde erkannt, daß für die Psychotherapieforschung die Kriterien zu wirklichkeitsfern sind. Stattdessen werden zunehmend naturalistische Studien und auch qualitative Forschungen in die Untersuchung der Wirksamkeitsnachweise (bzw. Evaluation) von Psychotherapie einbezogen.“

Es wäre unangemessen, diese Positionen so gewichtiger Organisationen im medizinischen Bereich, wie der SVR oder der AWMF, als Randerscheinungen abzutun. Stattdessen muss akzeptiert werden, dass die Frage nach Vor- und Nachteilen von EbM komplexer ausfällt, als es manche Befürworter aber auch Gegner gern hätten. Auf jeden Fall ist es aber auch wichtig darauf zu achten, dass EbM mit der guten Kernidee, Wissen und Entscheidungshilfen bereitzustellen, nicht für andere Zwecke missbraucht wird – was leider zunehmend zu beobachten ist. Selbst David Sackett (der „Vater“ der EbM) sah sich genötigt, kritisch vor "top-down cook-books", dem "misuse of evidence based medicine" zu warnen und darauf hinzuweisen, dass EbM allzu oft von „Kaufleuten“ und Managern dazu eingesetzt wird, einfach Kosten zu senken. Ebenso wird die Bedeutung von EbM durch Einrichtungen wie dem Gemeinsamen Bundesausschuss (G-BA) beschädigt, wenn diese – wie im aktuellen Fall des Ablehnungsverfahrens der Gesprächspsychotherapie – die EbM-Kriterien in völlig einseitiger und inadäquater Anwendung dafür missbrauchen, effektive Konkurrenz vom Markt fernzuhalten.

Fazit

Kern des Vortrags war es, dafür zu sensibilisieren, dass EbM einerseits eine sehr begrüßenswerte Kernidee beinhaltet, die aber andererseits umso schwieriger faktisch zu realisieren ist, wie die Zusammenhänge zwischen Interventionen und deren Auswirkungen an Komplexität zunehmen. Eine schematische Anwendung von EbM beispielsweise im Bereich von Psychotherapie, würde sich über eine erhebliche Zahl ungelöster Fragen und ungeprüfter Vorannahmen hinwegsetzen. Die eigentlich angestrebte Überwindung ungeprüfter Glaubenssätze durch eine fundierte Wissensbasis droht dann ins Gegenteil verkehrt zu werden, indem EbM selbst zu einer Glaubensgemeinschaft führt, die sich gegenüber Kritik und Wissen durch einen methodisch missverstandenen Rigorismus immunisiert. Daher müssen die vielen impliziten Vorannahmen kritisch auf ihre jeweilige Angemessenheit diskutiert und ggf. entsprechende Adjustierungen vorgenommen werden.¹¹

An einem Beispiel (hier im Anhang) wurde gezeigt, wie selbst dem scheinbar „einfachen“ Vergleich zweier Prozentwerte implizite Modellannahmen zugrunde liegen.

LITERATURHINWEISE

- ¹ vgl. Kienle, G. et al: Evidenzbasierte Medizin: Konkurs der ärztlichen Urteilskraft? Deutsches Ärzteblatt, 100, 2003
- ² Seligman, M.E. P. (1997): Die Effektivität von Psychotherapie. Die Consumer Reports-Studie, in: Integrative Therapie 22,4, 264-288 (amer. Original in: American Psychologist 50, 965-974
- ³ Kriz, J. (1996): Grundfragen der Forschungs- und Wissenschaftsmethodik. In: Hutterer-Krisch, R. et.al.: Psychotherapie als Wissenschaft - Fragen der Ethik, Bd. 5 der "Serie Psychotherapie" (Hrsg.: G. Sonneck), Wien: Facultas, 15-160
- ⁴ Kriz (2006): Wie evident ist klinische Evidenz? In: Bührmann, A. et al. (Hrsg.) Gesellschaftstheorie und die Heterogenität empirischer Sozial--forschung. Münster, WD-Verl. S. 32 - 43
- ⁵ Orlinsky, D. E. & Howard, K. (1987): A generic model of psychotherapy. J. Integrative Eclectic Psychother 6: 6-27. Deutsch: 1988. Ein allgemeines Psychotherapiemodell. Integrative Therapie 4: 281-308
- ⁶ Schmacke, N. (2006) Evidenzbasierte Medizin und Psychotherapie: die Frage nach den angemessenen Erkenntnismethoden *Psychotherapie, Psychosomatische Medizin und Medizinische Psychologie*, i.Dr.
- ⁷ SVR (1999): Bedarfsgerechtigkeit und Wirtschaftlichkeit, Bd. II: Qualitätsentwicklung in Medizin und Pflege, BT-Drs. 14/5661 v. 21.3.2001
- ⁸ Greenhalgh, T. (1999): Narrative based medicine in an evidence based world, in: BMJ Vol. 318, 30.1.1999
- ⁹ SVR (2003): Gutachten 2003. Kurzfassung (Quelle: Homepage des SVR)
- ¹⁰ Homepage der AWMF: www.awmf-online.de
- ¹¹ Verwiesen sei auch auf weitere Literatur des Autors auf: <http://jkriz.de/txtwien05.htm>

ANHANG

Die – oft impliziten – Modellannahmen bei selbst „einfachen Sachverhalten“ wurden an folgendem Beispiel dargestellt und diskutiert (zur Vermeidung ideologischer Verzerrung aus einem anderen Bereich als der Psychotherapie – aber leicht übertragbar):

Eine Firma produziere irgendein Gut **G**; für 2 aufeinander folgende Zeitperioden von je 4 Jahren wird zuerst Maßnahme(nbündel) A dann B ausprobiert. Es ergibt sich:

G (in irgendwelchen Einheiten):

$\frac{60}{I}$	$\frac{100}{I}$	$\frac{150}{I}$
A	B	

Argumentation 1: B besser als A, denn Zuwachs von 50 vs. 40 ist 25% mehr

Argumentation 2: B schlechter als A: Zuwachs von 50% (100→150) vs. 67% (60→100)= 17% weniger

wer hat Recht ?? (abgesehen davon, dass man üblicherweise nur Arg.1 oder Arg.2 liest!): **Beide !**

Arg.1 liegt implizit ein lineares Wachstumsmodell zugrunde (z.B. eine Maschine produziert G unter sonst gleichen Bedingungen). Zu erwarten wäre da für B nur 140, 150 ist also „wirklich“ besser.

Arg.2 liegt implizit ein dynamisch-autokatalytisches Wachstumsmodell zugrunde (z.B. Populationsdynamik: mehr G produziert auch mehr G) Zu erwarten wäre dann für B 67% mehr, also 167. Somit ist 150 also „wirklich“ schlechter.

Es fehlt – wie oft – die Information, welches implizite Modell angemessen ist (und die Entscheidung im je konkreten Fall ist in der Regel eine viel komplexere Debatte, als „Zahlengläubige“ vermuten)
(Beispiel nach Kriz, J. (1988): Facts and Artefacts in Social Science. Hamburg, NewYork: McGraw-Hill)