

Kritische Anmerkungen zur evidenzbasierten Psychiatrie¹

■ Daniel Hell

Psychiatrische Universitätsklinik, Zürich

Einleitung

Die Notwendigkeit von Qualitätsbeurteilungen ist unbestritten. Es ist höchst verdienstvoll, wenn Ärzte, Therapeuten und kritische Laien Zugang zu Studien haben, welche die Resultate einer Behandlungsmethode überprüft haben. Es ist ebenso verdienstvoll, wenn zum Beispiel in psychiatrischen Lehrbüchern [1] versucht wird, nicht nur Expertenmeinungen, sondern empirische Befunde zusammenzufassen.

Meine kritischen Bemerkungen richten sich nicht gegen die Qualitätsbeurteilung als solche oder gar gegen empirische Befunderhebungen, sondern fragen nach den Konsequenzen einer bestimmten Art von Qualitätsmessung und Evidenz für die heutige Psychiatrie und Psychotherapie. Auch Art und Umfang der Qualitäts- und Erfolgsmessung steht zur Diskussion, vor allem aber die langfristigen Folgen einer zu Vergleichszwecken nötigen Vereinheitlichung therapeutischer Ansätze.

Meine Ausführungen gehen von einer These aus und diskutieren die folgenden vier Punkte:

1. Zeitgeschichtlicher Hintergrund,
2. Wahrscheinlichkeit und Wahrheit,
3. Möglichkeiten und Grenzen der Statistik,
4. Gefahr der Ideologisierung von evidenzbasierter Medizin.

These

«Die Evidenzmessung ist ein wertvolles Instrument gegen Missbräuche und Scharlatanerie. Sie stellt nicht nur ein unverzichtbares Kontrollinstrument dar, sondern lässt auch (durch empirische Daten gestützte) Schlüsse zu, die sich abheben von blossen Lehrmeinungen oder anekdotischen Mitteilungen.

Sie generiert aber als Hilfsmittel keine eigenständige Medizin oder Psychiatrie und stellt keine gleichwertige Grundlage für die Psychiatrie wie z.B. Psychopathologie, Psychodynamik und Neurobiologie dar. Eine Überschätzung der Evidenzmessung als me-

thodisches Werkzeug leistet der Verarmung von Psychiatrie und Psychotherapie Vorschub.»

Eine Qualitätsbeurteilung, welche Teil der evidenzbasierten Medizin ist, hat es notwendigerweise mit Fragen der Praxis zu tun. Natürlich kann auch die Forschung selbst einer Qualitätsbeurteilung unterzogen werden. Dann wird aber ebenfalls ein praktisches Vorgehen einer kritischen Beurteilung unterzogen. Qualitätsforschung überprüft, aber sie findet keine Wahrheit, die jenseits des überprüften Sachverhaltes liegt. Sie dient dem Schutz der Forschenden und dem Schutz der Konsumenten. Das ist in einer Zeit der Globalisierung mit kaum mehr überschaubaren Verhältnissen und raschem medizinischem Fortschritt ausserordentlich wichtig. Aber Qualitätsforschung kann nicht zum Ziel haben, die Wirklichkeit über die untersuchten Studien hinaus abzubilden und Neues zu entdecken.

Warum muss dies betont werden? Der Begriff «evidenzbasierte Medizin» gibt schlagwortartig eine neue Medizin vor. Statt nur das Korrektiv zu betonen (wenn etwa gesagt würde: «praktische Medizin mit Qualitätskontrolle»), wird begrifflich zum Ausdruck gebracht, dass hier eine neue Medizin entsteht, nämlich eine, die auf Evidenz basiert. Evidenz sollte aber neudeutsch hier nur heissen: eine kontrollierte beziehungsweise nachweismessungsgestützte Medizin. Das ist im Grunde genommen aber eine Banalität. Voraussetzung jeder guten Medizin ist die Empirie und der Nachweis ihres Erfolges, also Forschung, bei der Begriffe wie Sensitivität, Spezifität und Reproduzierbarkeit grundlegend sind. Es stellt sich aber die Frage, ob die evidenzbasierte Medizin nicht mehr sein will als eine Medizin, die auch ihr Tun mit statistischen Methoden überprüft. Ich befürchte ja, doch darauf ist zurückzukommen.

Zeitgeschichtlicher Hintergrund

Zunächst erlaube ich mir einen historischen Exkurs: Meinem Vorgänger Klaus Ernst verdanke ich den Hinweis, dass statistische Methoden in der vorchristlichen Geschichte – schon vor den Pharaonen – zunächst zur Eintreibung von Steuern angewendet wurden. Deshalb, so schliesst Klaus Ernst, hafte der Statistik bis heute etwas Ungemütliches, ja Anrüchiges an.

War und ist bei den Steuern einfach zu bestimmen, wem die Statistik diene und

dient, ist die Frage bei der evidenzbasierten Medizin etwas komplizierter zu beantworten. Sie sollte in erster Linie den Benutzern, das heisst den Konsumenten, dienen. Kann sie aber auch zur Kontrolle des Benutzerverhaltens eingesetzt werden? Doch diese Frage will ich noch etwas zurückstellen.

Die zweite historische Bemerkung betrifft den Begriff «Evidenz». Er war zunächst ein gängiger philosophischer Ausdruck für etwas, das ohne weiteres Studieren klar ist. Wenn Husserl oder Heidegger den Begriff gebrauchten, beinhaltete er etwa das Gegenteil dessen, was er in der modernen Medizin meint: nämlich etwas, das wie ein Aha-Erlebnis nicht hinterfragbar ist, etwas Offensichtliches, eben Evidentes. Die Bedeutungskehr in der Moderne und der evidenzbasierten Medizin kann darauf hinweisen, dass das, was für das Erleben einer Einzelperson – aus der Perspektive der ersten Person – evident ist, heute in den Hintergrund getreten ist beziehungsweise nicht mehr «zählt» (auch im wahren Sinne des Wortes). Nur das, was mehrere Leute – aus der Drittpersonenperspektive – in gleicher Weise feststellen, scheint glaubwürdig zu sein. Die evidenzbasierte Medizin könnte denn auch als Zeichen unserer Zeit verstanden werden, die das Individuum offenbar zum verrechenbaren Objekt erklären muss, um es vor globalem Missbrauch zu schützen.

Die evidenzbasierte Medizin ist denn auch von G. E. Jackson [2] als postmodernes Phänomen charakterisiert worden. Tatsächlich kann sie als Versuch verstanden werden, in einer Welt, der ein einheitlicher Erzählstrang bzw. eine gemeinsame Wahrheit verloren ging, auf neue Weise Sicherheit zu geben. Neu ist auch, dass die Evidenz neuerdings auf Grund von Quantitäten, Populationen und Durchschnittszahlen statt wie früher auf Grund von Qualitäten, individuellen Berichten und spezifischen Beobachtungen bestimmt wird. Darin liegt ein Systemwechsel, der jedoch mehr mit dem philosophischen Zeitgeist des Positivismus, mit der Entwicklung von digitalen Technologien und von Biostatistik als mit einer vertieften Auseinandersetzung mit erkenntnistheoretischen Fragen zu tun hat.

¹ Vortrag an der Jahresversammlung der Schweizerischen Gesellschaft für Psychiatrie und Psychotherapie in Aarau am 2. September 2005, leicht gekürzte und veränderte Fassung.

Korrespondenz:

Prof. Dr. med. Daniel Hell
Psychiatrische Universitätsklinik
Lenggstrasse 31
Postfach 1931
CH-8032 Zürich
e-mail: hellsek@bli.unizh.ch

Wahrscheinlichkeit und Wahrheit

Frühe Vertreter einer evidenzbasierten Medizin haben zu Recht betont, dass auf statistischen Untersuchungen basierende Hinweise – die sogenannte externe Evidenz – die individuelle klinische Expertise – die sogenannte interne Evidenz – nur ergänzen kann. Sackett, ein prominenter Vertreter der evidenzbasierten Medizin, schreibt in einem vom deutschen Cochrane-Zentrum verbreiteten Basisartikel [3]: «Gute Ärzte nutzen sowohl klinische Expertisen als auch die beste verfügbare externe Evidenz, da keiner der beiden Faktoren allein ausreicht: Ohne klinische Erfahrung riskiert die ärztliche Praxis durch den blossen Rückgriff auf die Evidenz <tyrannisiert> zu werden, da selbst exzellente Forschungsergebnisse für den individuellen Patienten nicht anwendbar oder unpassend sein können. Andererseits kann ohne das Einbeziehen aktueller externer Evidenz die ärztliche Praxis zum Nachteil des Patienten leicht veraltetem Wissen folgen.»

Dem kann man nur uneingeschränkt zustimmen. Aber wie steht es denn mit den sich verbreitenden Behandlungsrichtlinien, die sich immer häufiger ausschliesslich auf externe Evidenz abstützen und z.B. dazu beitragen, dass nur bestimmte Behandlungsmethoden als Versorgungsstandard gelten und von den Krankenkassen bezahlt werden sollen? Hier scheinen statistische Methoden doch die individuelle Wahrheitsuche zu ersetzen statt zu ergänzen.

Wahrscheinlichkeit und Wahrheit haben aber nur den ersten Teil des Wortes gemeinsam. Was wahrscheinlich ist, *ist* nicht wahr, sondern *scheint* wahr zu sein. Es ist vielen Vertretern der evidenzbasierten Medizin zugute zu halten, dass sie die Aussagekraft statistischer Methoden hinterfragen und sich bemühen, passende Methoden für unterschiedliche Fragestellungen zu finden.

Um nochmals Sackett [3] zu zitieren: «Kliniker, die eine Kochbuchmedizin fürchten, werden sich jedenfalls mit den Advokaten der evidenzbasierten Medizin auf den Barrikaden wiederfinden.» Doch auch hier ist zu fragen, ob dieser 1997 geschriebene Satz von der real existierenden medizinischen Versorgung eingeholt worden ist. Oder habe ich die Barrikaden übersehen, auf denen Vertreter der evidenzbasierten Medizin zusammen mit praktizierenden Ärzten z.B. gegen die geplante Psychotherapieplafonierung gekämpft hätten?

Statistische Methoden können geeignet sein, die klinische Praxis zu überprüfen. Man denke etwa an die Behauptung, dass Antidepressiva Suizide verhindern. Jetzt haben aber statistische Analysen der FDA-Daten von 50 000 mit Antidepressiva oder Placebo behandelten Patienten und andere Studien gezeigt, dass in der akuten Erkrankungsphase, welche die ersten 6–7 Wochen umfasst, die Zahl der vollendeten Suizide unter SSRI doppelt so hoch und unter Trizyklika noch 1,5mal so hoch waren wie unter Placebo [4,5]. Diese durchaus weiter zu überprüfenden Daten können viele Ursachen haben, zum Beispiel eine an und für sich erwünschte Aktivitätssteigerung beziehungsweise eine

Verminderung der depressiven Blockade unter Antidepressiva. Doch weisen solche Befunde darauf hin, dass gängige Lehrmeinungen durchaus hinterfragbar sind. Es wäre aber auch verfehlt, sich nur auf grossangelegte statistische Untersuchungen zu verlassen, um im Einzelfall zu bestimmen, was richtig und was falsch ist. Statistische Daten können dem Arzt nur populationsorientierte Hinweise geben, die allerdings für seine Entscheidung hilfreich und handlungsrelevant sein können. Bekanntlich gelten Berechnungen eines Krankheitsrisikos, einer Erkrankungsprognose oder eines Behandlungserfolges nur für die untersuchten Stichproben – und wenn diese repräsentativ ausgewählt wurden – für eine Bevölkerung mit einer bestimmten Alters-, Geschlechts-, Schicht- und Kulturzugehörigkeit. Aber es ist natürlich irrig, diese statistische Durchschnittsberechnung so auf eine einzelne Person anzuwenden, als ob diese dem statistischen Bevölkerungsdurchschnitt entsprechen würde, als gäbe es den durchschnittlichen Schweizer mit einem durchschnittlichen Gehalt, einer durchschnittlichen Depression oder einer durchschnittlichen Persönlichkeitsstruktur. Diese Binsenwahrheit sei am Behandlungsverlauf unter Antidepressiva illustriert. In der Regel werden in Publikationen nur die Verlaufskurven des statistischen Durchschnitts einer Population (unter Antidepressiva und Placebo) dargestellt. Visuell wird gleichsam der Behandlungserfolg bei einem durchschnittlichen Patienten, den es nicht gibt, vorgestellt. Denn die interindividuelle Vielfalt ist enorm. Sie kommt erst zur Darstellung, wenn der unterschiedliche Verlauf jeder einzelnen behandelten Person graphisch erfasst wird.

Möglichkeiten und Grenzen der Statistik

Die einzelne Person ist keine Mittelung vieler Menschen, sondern ein Individuum, also eine unteilbare Person mit einer Selbstorganisation, die systemisch gesehen einer andern Ordnung angehört als eine Gruppe oder eine Gesellschaft. Was an einer Gruppe von Menschen gefunden wird – und sei es noch so methodisch perfekt –, lässt kein abschliessendes Urteil für einen einzelnen Menschen zu, sondern nur einen Schluss auf die untersuchte Population.

Es gibt zwar eindeutige Fälle, etwa der Sturz aus einem hoch fliegenden Flugzeug. Hier ist der Mensch noch am ehesten als ein mit allen andern Menschen vergleichbares Objekt zu behandeln, das den gleichen physikalischen Gesetzen, nämlich der Schwerkraft, unterliegt und durch den Fall zu Tode kommt – es sei denn, er sei mit einem Fallschirm ausgerüstet. Aber gerade solche eindeutigen Fälle lassen – wie Smith und Pell neulich mit englischem Humor in einem Artikel [6] konstatiert haben – keine kontrollierte und randomisierte Studie zu, weil niemand auf die Idee kommt, den Fallschirm durch ein Placebo zu ersetzen. Gleiches gilt für den suizidalen Menschen, der sich von einem Hochhaus oder vor einen Zug stürzen will.

Das andere Extrem stellen Situationen dar, die wie russisches Roulette dem Zufall unterliegen. Wo Zufall herrscht, lassen sich keine statistisch signifikanten Korrelationen zwischen einer Handlung und ihren Folgen finden. Für die evidenzbasierte Medizin sind jene Fälle am interessantesten, die in der Mitte zwischen diesen beiden Extremen liegen. Die Statistik seltener Ereignisse (wie z.B. vollendete Suizide) brauchen eine andere statistische Behandlung als häufige Ereignisse (wie z.B. depressive Erkrankungen). Die klinische Relevanz einer statistischen Signifikanz muss deshalb für jeden Fall gesondert betrachtet werden. Hier kann die evidenzbasierte Forschung aufklärend wirken und helfen, zwischen relevanten und irrelevanten Befunden zu unterscheiden. Denn: Was statistisch signifikant ist, muss klinisch noch lange nicht bedeutsam sein.

Vielleicht kennen Sie folgendes Beispiel: Die Wahrscheinlichkeit, dass eine Frau im Alter zwischen 40 und 50 Jahren, ohne Symptome, Brustkrebs hat, beträgt 1%. Wenn nun eine dieser Frauen Brustkrebs hat, dann beträgt die Wahrscheinlichkeit, dass sie einen positiven Mammographiebefund erhält, 80%. Wenn eine dieser Frauen jedoch keinen Brustkrebs hat, dann beträgt die Wahrscheinlichkeit, dass sie dennoch einen positiven Mammographiebefund erhält, 10%. Wie gross ist nun die Wahrscheinlichkeit, dass eine Frau mit einem positiven Mammographiebefund tatsächlich Brustkrebs hat?

Wenn man diese Fragen praktizierenden Ärzten stellt, antworten die meisten: 90%. Tatsächlich liegt das Risiko bei 7,5%. Denn zusätzlich zu den 8 positiv diagnostizierten Frauen mit Brustkrebs haben auch 10%, also 99 der 990 Frauen ohne Brustkrebs eine positive Mammographie, so dass die 8 Frauen mit Brustkrebs eben nur knapp 7,5% ausmachen.

Nicht unproblematisch ist auch die scharfe Trennung von Placebowirkung und Wirkung eines Verum-Präparates. Gewöhnlich wird davon ausgegangen, dass in placebokontrollierten, randomisierten Studien – der Goldstandard der evidenzbasierten Medizin – die Überlegenheit der aktiven Substanz gegenüber Placebo nachgewiesen werden kann. Es zeigt sich aber, dass an verschiedenen Kliniken vergleichbar durchgeführte Studien häufig zu widersprüchlichen Ergebnissen führen. Dazu kommt, dass das Verumpräparat spezifische Nebenwirkungen verursacht. An diesen ist nicht nur für den «geblindeten» Untersucher zum Teil erkennbar, um welche Behandlung es sich handelt, sondern es ist auch nicht ausschliessbar, dass die Nebenwirkung beim Patienten indirekt Einfluss auf die Wirkung hat. Nicht selten wird sogar die Nebenwirkung als Wirkung verstanden.

Dieser Effekt ist so lange nebensächlich, wie Verum- und Placebowirkung relativ weit auseinanderklaffen, zum Beispiel bei Antipsychotika. Wo diese beiden aber enger beieinander liegen, wie bei Antidepressiva, ist er nicht vernachlässigbar.

Meines Erachtens ist es beunruhigend, dass die Effektstärke der hauptsächlich angewendeten Antidepressiva in den letzten Jahrzehnten tendenziell abgenommen hat. Diese Beobachtung hängt sicher mit der Aus-

weitung der Indikation zur antidepressiven Behandlung auf Stimmungsschwankungen und der Zunahme von Vorbehandlungen durch niedergelassene Ärzte zusammen. Der Erfolg der antidepressiven Behandlung wird wesentlich durch therapeutische Gespräche verstärkt, doch dürfte die Bereitschaft zu solchen Gesprächen im letzten Jahrzehnt abgenommen und der isolierte Einsatz von antidepressiven Medikamenten zugenommen haben.

Gefahr der Ideologisierung von evidenzbasierter Medizin

Die grösste Gefahr der evidenzbasierten Psychiatrie scheint mir in der unkritischen Überbetonung statistischer Daten zu liegen.

Experimentelle medizinische Studien geschehen in einem artifiziellen Milieu. Sie können naturalistische Beobachtungen nicht ersetzen. Anders als Studien in der Physik führen angewandte psychiatrische und psychotherapeutische Forschungen nicht zur Erkennung von immer gültigen Naturgesetzen. Menschen treffen eigene Entscheidungen und verändern sich stetig. Auch die Anwendung komplexer mathematischer Methoden vermag dies nicht zu ändern. Das heisst nun nicht, dass statistische Verfahren in unserem Fach keine Anwendung finden sollen. Im Gegenteil. Es heisst aber, dass ihr Anwendungsgebiet genau definiert und ihre Interpretationsweise kritisch hinterfragt werden soll. Arithmetik, mithin Quantität, geschieht immer ohne «Ansehen der Person». «Zahlen stinken nicht», könnte man in Anlehnung an die bekannte Formel des Kaisers Vespasian «Geld stinkt nicht» sagen. Die Zahlenarithmetik entkleidet alle betroffenen Personen ihrer persönlichen Qualitäten als Einzelmenschen. Der Mechanismus der zählenden Abstraktion muss so wirken. Nur so kann er auch in der Medizin und Psychiatrie seine Vorzüge zur Geltung bringen.

Die Frage ist nur, wer zu welchem Zweck welche Zahlen zusammenstellt und wer sie interpretiert. Letztlich entscheidet das Referenzsystem über die Auswahl und die Interpretation der Daten. Aber ein Referenzsystem ist nun einmal durch die beste Computerwissenschaft nicht selber zu generieren.

Eine Bewertung von Qualität bzw. Evidenz ist ohne Referenzsystem nicht denkbar. Qualität, lat. *qualitas*, meint nämlich immer nur ein Sosein, eine Art und Weise, die Beschaffenheit eines Gegenstandes oder eines Handelns [7]. Erst in einem grösseren Sinnzusammenhang ist zu beurteilen, ob eine bestimmte Art und Weise, eine bestimmte Qualität eben, gut oder schlecht ist.

Sind kurze Hospitalisationszeiten grundsätzlich wünschenswert?

Sind niedrige Medikamentendosierungen immer positiv zu bewerten?

Sind häufige Gespräche ein positives Qualitätsmerkmal?

Die Qualität eines Angebots oder die Evidenz eines Handelns in der Psychiatrie misst sich letztlich an unseren Erwartungen, also am Referenzsystem. Aber was ist mit *uns* gemeint?

Ich denke mir, dass es vergleichsweise einfach ist, von Person zu Person übereinstimmende Qualitätsanforderungen für die Wohnqualität innerhalb und ausserhalb psychiatrischer Institutionen zu finden. Kaum jemand wird daran zweifeln, dass es für psychisch Kranke günstiger ist, über ein individuell eingerichtetes Zimmer zu verfügen, als langfristig in Bettensälen ohne private Einrichtungsgegenstände leben zu müssen. Auch die Definition von Nebenwirkungen psychopharmakologischer Behandlungen wird kaum Anlass zu Diskussionen geben. Wer aber bestimmt Qualitätsanforderungen an die therapeutische Beziehung oder an zwischenmenschliche Betreuungsformen? Wer entscheidet, ob eine vorübergehende Verschlechterung zugunsten eines längerfristigen Behandlungserfolges in Kauf genommen werden darf? Wer misst Trauer gegen Depression?

Ich erhoffe mir, dass sich die evidenzbasierte Psychiatrie auf dem Weg zur computergestützten Sicherung der Qualität zu entscheiden weiss, was messbar ist und was nicht bzw. wo die Grenzen der Messgenauigkeit liegen. Auch hoffe ich, dass in unserem Fach trotz allem technisch Machbaren und trotz zunehmender Zahlenflut die unwägbareren Bereiche nicht übersehen oder vergessen werden.

Eines bin ich mir recht sicher. Statistische Analysen schaffen keine Welt, die sich be-

wohnen lässt. Randomisierte und kontrollierte Untersuchungen konfigurieren eine Datenmenge. Diese virtuelle Realität lässt wie ein Flugsimulator gewisse Trainingsübungen durchführen. Fliegen beziehungsweise Behandeln in der Praxis bedeutet aber immer noch einen Sprung aus der virtuellen Realität in die existentielle Wirklichkeit. Es wäre fatal, virtuelle und existentielle Realität miteinander zu verwechseln, so hilfreich digitale Instrumente auch sind.

Literatur

- 1 Berger M, Herausgeber. Psychische Erkrankungen. München: Urban Fischer; 2004.
- 2 Jackson GE. EBM – A postmodern phenomenon [letter to the editor]. Br Med J 22nd February 2004.
- 3 Sackett LE, Rosenberg W, Muir-Gray JA, Haynes RM, Richardson WS. Was ist Evidenzbasierte Medizin und was nicht? Münch Med Wschr 1997;139:644–5.
- 4 Khan A, et al. Placebo-Treated Depressed Patients in Antidepressant Clinical Trials: Suicide Risk and Magnitude of Symptom Reduction. 4th Annual Meeting of New Clinical Drug Evaluation Unit NIMH, 2000.
- 5 Fergusson D, Doucette S, Glass KC, Shapiro S, Healy D, Hebert P, et al. Association between suicide attempts and selective reuptake inhibitors: systematic review of randomised controlled trials. Br Med J 2005;330:396–9.
- 6 Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. Br Med J 2003;327:1459–61.
- 7 Hell D. Qualität oder Güte? In: Hell D, Bengel J, Kirsten-Krüger M, Herausgeber. Qualitätssicherung der psychiatrischen Versorgung. Basel: Karger; 1998.