

Kritik der reinen Evidenz

Homöopathie in der evidenzbasierten Medizin – Teil 1

Georg Ivanovas
Homöopathie **KONKRET** 3.9: S. 10-18.

Zusammenfassung

Die evidenzbasierte Medizin (EBM) gewinnt in der medizinischen Praxis ständig an Bedeutung. Die meisten überzeugten Vertreter dieser Methode sehen in der Homöopathie nichts anderes als eine Placebothherapie. Dies spiegelt sich insbesondere in den großen medizinischen Zeitschriften wider. Diese Wahrnehmung stellt für die Homöopathie in vieler Hinsicht eine Herausforderung und ein Problem dar.

Dieser Artikel befasst sich mit der Begrenztheit der evidenzbasierten Methode im Allgemeinen. Er zeigt an detaillierten Beispielen, wie ihre Techniken nicht oder nur unzureichend geeignet sind, Heilungsverläufe angemessen zu beurteilen. Darüber hinaus wird gezeigt, dass die real-existierende EBM nicht die best mögliche Evidenz bereitstellt. Sie ist zu anfällig für die verschiedensten Interessen, was derzeit auch zu Lasten der Homöopathie geht.

Die so genannte Evidenz basierte Medizin (EBM) gilt inzwischen allgemein als der Goldstandard in der Medizin. Dabei handelt es sich um eine Medizin, die sich ganz auf randomisierte, doppelblind kontrollierte Studien stützt. Passionierte Vertreter dieses Verfahrens fordern, dass allen, die sich nicht daran halten, die Therapieerlaubnis entzogen werden sollte (52). Der Therapeut sollte quasi „ferngesteuert werden durch beispielsweise Epidemiologen, Statistiker, Staat und Krankenkassen“ (31). Auch wenn die meisten Entscheidungsträger nicht ganz so weit gehen, so wird dem statistischen Nachweis immer mehr Bedeutung gegeben.

Diese Situation ist insofern prekär, als die Homöopathie in Untersuchungen der evidenzbasierten Medizin nicht besonders gut abschneidet. Sie kommt, wenn es um den Wirksamkeitsnachweis geht – und es geht immer häufiger darum – immer weiter in die Defensive. Dieses Thema ist für die rechtliche Situation der Homöopathie in den kommenden Jahren und Jahrzehnten absolut entscheidend. Es ist deshalb unerlässlich, die Thematik genau zu kennen und die Stärken, Schwächen und Begrenzungen der evidenzbasierten Medizin einschätzen zu können.

Wer heilt hat nicht notwendigerweise Unrecht

Wenn Patienten mit einer Plastiktüte voller Medikamente – alle auf ihre Evidenz geprüft – zu uns kom-

men und trotz deren Einnahme seit Jahren oder Jahrzehnten krank sind, dann scheint das irgendwie Schicksal zu sein, an der Genetik zu liegen oder an der Umwelt. Wenn dieselben Patienten nach 1–2 Jahren homöopathischer Therapie keine Medikamente mehr nehmen und gesund sind, dann ist das zwar kein Beweis für die Wirksamkeit der homöopathischen Therapie, es ist aber ein Beweis dafür, dass mit dem Wirksamkeitsnachweis irgendetwas nicht ganz stimmt. Es gibt also ein Problem mit unserem Goldstandard, auch wenn dessen Vertreter dies nicht so gerne hören oder akzeptieren wollen.

Relevanz für das Individuum

Randomisierte Studien können sehr nützlich sein. Bereits 1862 konnte in vergleichenden Studien nachgewiesen werden, dass der Aderlass bei Patienten mit Pneumonie keine wirksame Therapie ist (7).¹ Vergleichsweise können viele Praktiken als unsinnig oder schädlich erkannt werden. So zeigte die CRASH-Studie, dass die Gabe von Kortikosteroiden nach Schädelhirntrauma (damals die Standardtherapie) zu einer schlechteren Prognose führt (12).

Doch nicht immer liegen die Dinge so einfach. Beispielsweise zeigen randomisierte Studien, dass die Gabe von Antibiotika bei der Otitis media keinen wesentlich besseren Effekt aufweist als die Gabe eines Placebos, dafür häufiger zu Rezidiven führt (26; 50). Dies führte konsequenterweise zu einer Änderung der offiziellen Therapie-Empfehlung bei Otitis media. Sie besagt, dass man zumindest 48 Stunden abwarten soll, bevor eine antibiotische Therapie begonnen wird. Al-

lenfalls soll schmerzlindernd/antipyretisch behandelt werden. Obgleich das sicher keine schlechte Empfehlung ist, weist sie bereits auf eines der vielen Probleme der EBM hin („in a nutshell“, wie es im Angelsächsischen so schön heißt). Zunächst einmal gibt es für die Effektivität der antipyretischen Therapie keine Evidenz, dafür aber Hinweise für ihren negativen Einfluss (5). Zum anderen stellt sich aber die Frage, warum eigentlich nach 48 Stunden mit der Antibiose begonnen werden soll. Medizinisch ist es offensichtlich: Es gibt bakterielle Otitiden, die, wenn das Immunsystem nicht mit ihnen klar kommt, zu extremen Verläufen neigen (Mastoiditis, Meningitis). Solche Fälle sollte man tunlichst rechtzeitig antibiotisch behandeln, wenn andere Methoden nichts nützen oder nicht zur Verfügung stehen. Das theoretisch-statistische Problem besteht darin, dass laut EBM die Antibiose praktisch „ineffektiv“ ist, was natürlich nicht stimmt. Sie ist nur ineffektiv, wenn alle Otitiden in einen Korb geworfen werden, wenn also nach einem sehr groben klinischen Kriterium randomisiert wird.

Das Problem ist, dass sich nicht vorhersagen lässt, welche Kinder mit welcher Symptomatik profitieren werden. Wenn man versucht, nach einer mehr differenzierten Symptomatik zu randomisieren, entstehen zu kleine Gruppen, die statistisch zu unzuverlässigen Aussagen führen (50).

Renommiertere Kritiker der statistischen Medizin haben dieses Problem so formuliert: „Große Zahlen liefern ein statistisch gesehen genaues Ergebnis, von dem man nicht weiß, auf wen es zutrifft. Kleine Zahlen liefern ein statistisch gesehen unbrauchbares Ergebnis, von dem man aber besser weiß, auf wen es zutrifft. Schwer zu entscheiden, welche dieser Arten von Unwissen die nutzlosere ist.“ (6).

Fast identisch argumentierte übrigens auch schon Bernard ganz zu Beginn unserer heutigen wissenschaftlichen Medizin.²

Dazu noch folgendes Beispiel (43):

- Die ALLHAT-Studie bestimmt die effektivsten Medikamente bei Hochdruck, gibt aber keinen Hinweis darauf, für welchen Patienten welche Darreichung und Dosierung die beste ist.
- Die MADIT-Studie kommt zu klaren Aussagen über kaukasische Männer zwischen 55 und 75 Jahren mit koronarer Herzkrankheit, komplexen ven-

trikulären Arrhythmien und implantiertem Defibrillator. Aber wie viele solcher Patienten haben wir schon?

Einen anderen Aspekt desselben Problems liefert eine Studie über die Härte von Matratzen. Es wurde eine randomisierte Doppelblindstudie zum Einfluss der Matratzenhärte auf den Rückenschmerz (low back pain) durchgeführt. Das mag auf den ersten Blick unbedeutend erscheinen. Aber wer mit dem Thema vertraut ist, weiß, dass die Matratzenhärte ein zentrales Problem bei Patienten mit chronischen Rückenschmerzen darstellt. Die Matratze ist oft ausschlaggebend für das Wohlergehen des Patienten, für seine Arbeitsfähigkeit und für das gesamte Familienklima. Kovacs et al. führten eine randomisierte, doppelblinde Studie über 90 Tage durch. Das ist zwar etwas kurz, aber es geht hier nicht um das Ergebnis, sondern um die Problematik der Interpretation. Die Studie fand heraus, dass mittelharte Matratzen ein besseres Ergebnis für Schmerz und Bewegungsfähigkeit aufwiesen als harte Matratzen (33). So weit, so gut. Aber 10 % der auf mittelharten Matratzen schlafenden Patienten hatten mehr Schmerzen. Sollte man also allen empfehlen, auf mittelharten Matratzen zu schlafen, weil es für das Kollektiv besser ist? Oder sollte die Reaktion des Patienten den Ausschlag geben? Wenn wir nach der Reaktion gehen, was eigentlich sinnvoll wäre, dann spielen plötzlich all jene unspezifischen Faktoren eine Rolle, die Placebo-Effekt genannt werden (siehe HK 2.09), und die ja gerade durch die randomisierten Studien eliminiert werden sollen.

Randomisierte Studien können keine Aussagen darüber machen, wer von einer Therapie profitieren wird und wer nicht, haben keine Relevanz für die individuelle Prognose. Die so genannte NNT (number needed to treat – Anzahl der zu behandelnden Patienten) sagt aus, wie viele Patienten behandelt werden müssen, so dass einer davon profitiert. NNTs unter 5 sind ungewöhnlich, während Zahlen über 20 durchaus normal sind (59).

Die so genannte „maßgeschneiderte Medizin“ (tailored medicine), von der heute immer mehr geredet wird, kann nicht auf der Statistik beruhen, sondern nur auf der individuellen Beobachtung und auf einem Konzept einer individuellen Diagnose. Ein solches Konzept existiert aber nicht in der orthodoxen Medizin und erst recht nicht in einer evidenzbasierten Me-

dizin, die individuelle Verläufe statistisch sozusagen „einebnen“.

Hilft kurzfristig, schadet langfristig

Eine weitere wesentliche Limitation der evidenzbasierten Medizin ist die Dauer der Studien. Nehmen wir folgendes Beispiel: In einer randomisierten Studie über Epicondylitis wurden 3 Therapiegruppen gebildet. Die Patienten der Gruppe 1 erhielten eine Kortison-Injektion. Die Patienten der Gruppe 2 wurden krankengymnastisch behandelt. Gruppe 3 war eine Kontrollgruppe ohne besondere Therapie. Das Ergebnis war eindeutig: Nach kurzer Zeit waren die Patienten in der Kortison-Gruppe deutlich gebessert, während es in den anderen Gruppen nur zu geringen oder keinen Veränderungen kam. Nach einem Jahr jedoch hatten die Patienten der Kortison-Gruppe deutlich mehr Beschwerden als jene, die krankengymnastisch oder gar nicht behandelt wurden (58). Da das Ergebnis sehr unwahrscheinlich schien, wurde die Studie von einer anderen Gruppe mit demselben Ergebnis wiederholt (8). Sollte das Prinzip „hilft kurzfristig – schadet langfristig“ weiter verbreitet sein, dann wäre die Evidenz aller Studien von kurzer Dauer (also der meisten) geradezu eine Anleitung zur Schädigung des Patienten.

Es gibt einige gewichtige theoretische und praktische Erwägungen, die dieses Prinzip nahe legen. Hier aber zunächst einmal ein paar willkürliche Beispiele:

1. Eine Studie zeigte, dass Lauflernhilfen Kinder schneller in eine aufrechte Position bringen, das eigenständige Laufen jedoch verzögern und zwar umso länger, je länger die Lauflernhilfen verwendet wurden (22).
2. Säuglinge unter 6 Monaten, die Antibiotika erhielten, hatten eine höhere Wahrscheinlichkeit Allergien oder Asthma zu entwickeln (27).
3. Frühchen, die eine angereicherte Nahrung erhielten, entwickelten in einem höheren Prozentsatz eine Insulin-Resistenz in der Pubertät als jene, die eine nieder-kalorische Ernährung, Muttermilch oder eine Standard-Säuglingsmilch erhielten (56; 57). Das bedeutet, dass ein schnelles Anfütern nicht so gesund

ist wie eine langsame Entwicklung, selbst wenn die Normwerte früher erreicht werden.

4. Gestillte Säuglinge weisen höhere Gesamtcholesterin- und LDL-Werte auf als flaschengefütterte. Nach der Pubertät liegen diese Werte bei ihnen jedoch niedriger als bei den flaschengefütterten (45).

Die letzten beiden Studien liegen weit jenseits des üblichen Beobachtungszeitraums.

Wenn man die Effektivität einer Therapie über einen langen Zeitraum betrachtet (und es gibt keinen vernünftigen Grund, dies nicht zu tun), kann sich ein „positiver“ Effekt geradezu umkehren.

In den 1980er Jahren bestand die Standardtherapie der Osteoporose in der Gabe von Natriumfluorid, das die Aktivität der Osteoblasten anregt. Fast jede Woche bekamen wir Praktiker eine neue (meist industriegesponserte) Studie auf den Tisch, die die Wirksamkeit dieser Therapie eindrücklich belegte. Erst eine lang dauernde und unabhängige Studie zeigte, dass Natriumfluorid in der Tat die Knochendichte der Behandelten erhöhte, den Knochen aber gleichzeitig brüchiger machte (48).

Am besten untersucht in dieser Hinsicht wurde die Hormonersatztherapie mit Östrogen/Gestagen (HRT) bei Frauen (siehe Links). Die WHI-Studie verfolgte die gesundheitliche Entwicklung von mehr als 16.000 Frauen über einen Zeitraum von 5 Jahren. Dann wurde sie abgebrochen, da sie eine höhere Gefährdung in der Therapie-Gruppe feststellte. Bei den mit Hormonen Behandelten fanden sich mehr Fälle von Brustkrebs, Apoplexen und Thrombosen (11) und eine höhere Zahl frisch diagnostizierten Astmas (4). Vergleichbare Ergebnisse wurden auch bei einer Östrogen-Monotherapie beobachtet (64). Dies war kein zufälliges, statistisches Phänomen. Nach der Publikation der WHI-Studie kam es zu einer erheblichen Reduktion der Hormontherapien und, in der Folgezeit, zu einer deutlichen Reduktion der Brustkrebs-Diagnosen (47).

Bis zur WHI-Studie gab es eigentlich nur positive Wirksamkeitsstudien zur HRT. Noch kurz vor den ersten Veröffentlichungen zur WHI schrieb das *New England Journal of Medicine* geradezu euphorisch: „Die Entwicklung oraler Antikonzeptiva stellt einen wesentlichen Fortschritt für die Gesundheit der Frauen im letzten Jahrhundert dar“⁴³ (44). Dieses Beispiel

zeigt: Trotz Hunderter positiver Studien und Berichte in renommierten Zeitschriften besteht keinerlei Gewähr, dass eine Therapie wirklich nützlich ist.

Dies ist nicht auf die Hormonersatztherapie beschränkt. Aufsehen erregte der Fall von Rofecoxib (Vioxx), eines nichtsteroidalen Antiphlogistikums, das zu einem der größten Arzneimittelrückrufe führte. Die Zulassung von Rofecoxib erfolgte aufgrund positiver Studien mit einer Dauer von 3–6 Monaten. Danach wurde es aggressiv vermarktet. Etwa 80 Millionen Patienten wurden damit behandelt, was einen jährlichen Umsatz von etwa 2,5 Milliarden US \$ bedeutete. Längerfristige Studien zeigten später, dass Rofecoxib und vergleichbare Stoffe keinen Vorteil gegenüber anderen nichtsteroidalen Antiphlogistika hatten, dafür jedoch zu vermehrten kardiovaskulären Komplikationen wie Herzinfarkt (16/1000 Patienten) führten (20; 62). Ausgewählte, kurze Studien verbunden mit einer aggressiven Vermarktung erzeugen ein falsches Bild von Wirksamkeit. Eine vergleichbare Strategie wird derzeit gegen Homöopathie verwendet.

Etwas anders lag der Fall bei der Behandlung des Diabetes mellitus Typ 2. Stets wurde – aufgrund eher kürzerer Studien – eine „optimale“ (das heißt enge) Einstellung des Blutzuckers angestrebt. Es zeigte sich jedoch in längerfristigen Studien, dass dies zu vermehrten Todesfällen führt (53). Das ist nun nicht ganz so leicht zu verstehen und führt in einen Bereich, für den es in der universitären Medizin kein detailliertes Modell gibt: das Thema der Robustheit.

Robustheit und Rigidität

Robustheit ist ein Begriff, der vor allem in den traditionellen Naturheilverfahren verwendet wird. In der modernen Psychotherapie spricht man neuerdings von Resilienz, was in etwa dasselbe meint. Es geht dabei um die Fähigkeit, in bestimmten Situationen gesund zu bleiben oder zu überleben. Ein klassisches Beispiel wäre die Frage, warum manche der Häftlinge aus Konzentrationslagern später fröhliche Menschen wurden und andere den Rest ihres Lebens gebrochen waren. Warum werden manche Kinder aus zerrütteten Familien drogenabhängig oder straffällig und andere nicht? Warum ist für Manche eine Scheidung

oder der Tod des Partners zerstörend und für andere nicht? Die Erforschung der Faktoren, die robust bzw. resilient machen ist derzeit ein „heißes Thema“ in der Psychotherapie.

In der Allgemeinmedizin wird diesem Thema keine Bedeutung beigemessen. Dazu soll ein aktuelles Beispiel dienen: Bleuler schrieb nach der spanischen Grippe 1918/1919, dass er das Verhalten der Ärzte als eine Schande betrachte, da es versäumt wurde zu verstehen, was Patienten so robust macht, dass sie eine solche Epidemie unbeschadet überstehen. Er verwies darauf, dass es in seiner Klinik zu keinen Todesfällen gekommen sei, während vergleichbare Institutionen hohe Todeszahlen meldeten. Er führte dies darauf zurück, dass er die Korridore seines Krankenhauses nie und die Krankenzimmer nur wenig heizte (9). In der orthodoxen Medizin wird selten erörtert, was die Widerstandskraft erhöht. Ausgenommen davon sind die Impfungen. Vielleicht werden sie deshalb so hoch eingeschätzt, da es sonst kaum Strategien zur Verbesserung der Robustheit gibt.

So ergibt eine Suche in *pubmed* unter den Stichwörtern „H1N1“ und „recommendations“ (31.7.09) nur Artikel über Impfungen und Medikamente (Neuraminidase-Hemmer). Neuraminidase-Hemmer werden zwar intensiv diskutiert, obgleich deren Wirksamkeit eher moderat ist, wenn sie überhaupt wirken (31). Andere Strategien zur Verbesserung der individuellen Immunreaktion sind nicht aufzufinden. Die Suche unter den Stichwörtern „H1N1“ und „diet“ führt zu einem älteren Artikel, der (aufgrund von Tierexperimenten) eine Energierestriktion empfiehlt (48). „H1N1“ und „resilience“ ergibt keinen Treffer. In den Treffern mit den Suchwörtern „H1N1“ und „robustness“ geht es um die Robustheit der Daten. Seit Bleulers Zeit scheint sich also nichts geändert zu haben. Dabei ist ein differenziertes Verständnis von Robustheit unerlässlich, um die Aussagekraft von Wirksamkeitsstudien richtig einzuschätzen zu können, wie die nachfolgenden Beispiele zeigen sollen.

Nehmen wir das Beispiel von Betablockern. Diese zeigen eine sehr positive Wirkung bei den verschiedensten Erkrankungen – von der Hypertonie bis zum Glaukom. Aber sie schränken zu einem gewissen Grad die freie Regulation ein. Dies wird in der Homöopathie Unterdrückung genannt, wobei der homöopathi-

sche Begriff uneinheitlich verwendet wird und seiner wissenschaftlichen Definition noch harrt.

Personen, die mit Betablockern behandelt werden, haben eine schlechtere Wärmeregulation und erleiden leichter ein Hyperthermiesyndrom, was ein unterdiagnostizierter Notfall ist (10). So könnte eventuell bereits eine Hitzeperiode all den Überlebensvorteil zunichte machen, den die Studien zur Hypertonie usw. ergaben. So starben während einer Hitzewelle in Frankreich zwischen dem 4. und 18. August 2003 etwa 15.000 Menschen (46). Wie viele als Folge einer medikamenteninduzierten Rigidität? Normalbedingungen geben keinen Aufschluss darüber, wie sich ein Organismus in extremen Situationen verhält. Therapien, die in einem reizarmen Kontext positive Ergebnisse aufweisen, können sich in fordernden Situationen als fatal erweisen.

Die „Randomised Aldactone Evaluation Study“ (RALES) zeigte, dass Patienten mit schwerer Herzinsuffizienz durch Spironolacton ein deutlich verbessertes Ergebnis aufwiesen. Die Veröffentlichung der RALES-Studie führte nun zu einer vermehrten Verschreibung von Spironolacton. Die Folge waren eine deutlich erhöhte Morbidität und Mortalität durch hyperkaliämische Zustände (28). Die Autoren dieser Folgestudie und das begleitende Editorial (38) kamen zu dem Schluss, dass die Praktiker Fehler gemacht hätten, indem sie die Risikofaktoren nicht richtig eingeschätzt hätten und/oder die Kaliumwerte nicht häufig genug kontrolliert hätten. Das mag in dieser Form durchaus stimmen. In der Praxis werden immer (manchmal sogar bevorzugt) Patienten behandelt, die aus Studien ausgeschlossen wurden, da sie erwartungsgemäß zu einem weniger guten Ergebnis geführt hätten. Eine andere, sicher genauso stichhaltige Erklärung ist aber, dass unter den Bedingungen der Allgemeinpraxis Umstände eintreten, die die Grenzen der Robustheit einer Person überschreiten, also Umstände, die unter den kontrollierten Verhältnissen einer Studie niemals vorkommen. Die Aussagekraft von Studien ist erheblich einschränkt, wenn sie den Einfluss einer Therapie auf die Robustheit des Patienten nicht hinreichend in Betracht zieht. Eine Therapie, deren Ziel die Erhöhung der Robustheit ist (wie beispielsweise eine Kneippsche Therapie) mag in einer Studie nur einen geringen Effekt zeigen und einer medikamentösen Therapie, die den Patienten in die Rigidität zwingt,

deutlich unterlegen sein. Mittel- und langfristig könnten sie aber die sehr viel wirkungsvollere Therapie sein.

Der menschliche Faktor

Irren ist menschlich. So kommt es in wissenschaftlichen Veröffentlichungen immer wieder zu Fehlern. Eine Untersuchung von Artikeln der Zeitschrift *Nature* und dem *British Medical Journal* zeigte, dass etwa ein Drittel aller Artikel Fehler aufwies, wobei die Signifikanz der Ergebnisse in 12 % der Fälle veränderte wurde (21). Da es sich hierbei um renommierte Zeitschriften mit höchstem Standard handelt, bei denen alle Artikel von Statistikern kontrolliert werden (1), ist zu erwarten, dass der Anteil der Fehler in anderen Zeitschriften erheblich höher liegt. Aber obgleich dieser Prozentsatz schon recht hoch ist, stellt der Irrtum nicht das eigentliche „menschliche Problem“ der EBM dar.

Betrügen ist auch menschlich. Es konnte zwar auch bei einer Vielzahl von Tierarten nachgewiesen werden, aber das Ausmaß des Betrugs im Wissenschaftsbereich unterscheidet sich grundlegend von allen Beobachtungen im Tierreich.

Nature schätzt, dass ein Drittel aller wissenschaftlicher Artikel manipuliert (fiddled) ist (36), wobei andere Schätzungen – basierend auf Kollegeneinschätzungen – noch höhere Prozentzahlen annehmen (19). Gelegentlich werden Studien auch einfach erfunden (23; 14).

Daneben gibt es eine Vielzahl von Tricks, Zahlen zu schönen, ohne wirklich zu fälschen (17). Nun hört man immer wieder den augenzwinkernden Anspruch, dass jeder sich seine Statistik selber fälsche. Aber solche Totschlag-Argumente sind nur ärgerlich, da sie eine Beleidigung für all diejenigen sind, die sich um eine saubere Medizin bemühen.

Ein weiteres großes Problem neben der Fälschung/Schönung ist die Unterschlagung von Studien, vor allem durch die Pharmaindustrie. So lagen sowohl bei der Hormonersatztherapie (39; 40) als auch bei Rofecoxib (Vioxx) (2) den Pharmafirmen unerwünschte Zahlen vor, die sie nicht veröffentlichten. Im Fall der WHI wurden in dieser Zeit sogar noch positive Studien von Ghostwritern verfasst, die dann unter dem Namen renommierter Autoren erschienen (55).⁴ Die betroffenen Pharmafirmen versichern natürlich, dass das Datenmaterial dieser Veröffentlichungen korrekt

sei. Aber selbst wenn dem so wäre, so werden unerwünschte Ergebnisse selten veröffentlicht (41). Gelegentlich werden Wissenschaftler sogar regelrecht unter Druck gesetzt, um solche Veröffentlichungen zu verhindern (3; 51).

Jeder ist frei, sich vorzustellen, wie groß der Eisberg unter solchen Spitzen ist. Der ehemalige Herausgeber des *British Medical Journals* nannte die medizinischen Zeitschriften sogar eine „Verlängerung des Vermarktungsarmes der pharmazeutischen Gesellschaften“ (60)⁵. Um diesem Missbrauch zu begegnen, schufen die medizinischen Zeitschriften ein internationales Komitee, bei dem alle Studien vor deren Beginn angemeldet werden müssen (15). Das ist gut gemeint, funktioniert aber auch 5 Jahre nach seiner Etablierung nicht wirklich (37).

Pharmazeutische Firmen bestechen nicht nur Ärzte (25). Sie beeinflussen die öffentliche Wahrnehmung auf allen Ebenen, von den politischen Entscheidungsträgern, über die Veröffentlichung in Zeitschriften, bis hin zu den Selbsthilfeorganisationen der Betroffenen (24). Eine öffentliche Diskussion findet kaum statt. Dieselben Leute sprechen in Konferenzen über dieselben Themen, oft nur mit Erlaubnis von Pharmafirmen (63)⁶, so dass wir weit von einer wirklichen Evidenz entfernt sind. Aus all diesem, oft sehr zweifelhaften, Datenmaterial werden dann Übersichten erstellt, so genannte **Meta-Analysen**, die als höchster Grad der Evidenz gelten. Zumindest in der Theorie.

Nein, ein Goldstandard ist die real-existierende evidenzbasierte Medizin nicht.

Evidence Biased Medicine⁷ – parteiische Medizin

Ein weiteres Problem ist, wie mit den statistischen Daten umgegangen wird. Es ist ja nicht so, dass Statistiken etwas beweisen würden. Sie liefern Korrelationen. Die Bedeutung dieser Korrelationen wird durch deren Interpretation erzeugt. Das führt zu dem sehr heiklen Thema der Kausalität. Man kann zwar annehmen, dass eine Beobachtung A für eine Beobachtung B determinierend ist. Aber das ist nicht eine Frage von Realität, sondern eine Frage der Interpretation. Kausalität ist stets die Theorie eines Beobachters (siehe Links). Das klassische Beispiel der Statistik ist der

Rückgang der Störche, der deutlich mit dem Rückgang der Geburten korreliert. Wenn wir also von der Theorie ausgehen, dass der Storch die Kinder bringt, dann könnten wir in der Tat eine kausale Beziehung zwischen den beiden Ereignissen annehmen.

Medizinisch anschaulicher ist die Caerphilly Studie (18). Diese Studie fand eine erhöhte Anzahl von Apoplexien (70 %) und eine generell erhöhte Sterblichkeit (30 %) bei den Männern, die sich seltener als ein Mal pro Tag rasierten. Es war den Autoren nicht möglich, die erhöhte Mortalität mit anderen Faktoren überzeugend zu korrelieren. Rasieren war das Entscheidende. Da es sich um eine relativ große Studie handelte (2.438 Männer wurden über mehr als 20 Jahre verfolgt), müssen die Ergebnisse ernst genommen werden. Hier sind einige der Konsequenzen, die aus ihr gezogen werden sollten, könnten oder müssten:

- Aufklärungskampagnen über den Nutzen des Rasierens,
- Verordnungsfähigkeit von Rasiermitteln,
- Studien über den Einfluss von Depilationscremes auf Apoplex,
- Studien über den Zusammenhang zwischen Bartlänge und Sterblichkeit.

Hier ist sie wieder, die Realsatire des medizinischen Alltagsbetriebs. Sie fällt nur nicht so sehr auf, da es sich meist um medikamentöse Interventionen bei Diabetes, rheumatoider Arthritis oder Hypertonie handelt.

Zu glauben, dass Statistiken etwas beweisen, ist wie Ostereier suchen, die man selber versteckt hat.

Gehen wir zurück zur Homöopathie. Als eine erste große Meta-Analyse zur homöopathischen Therapie im *Lancet* erschien (35), die ihr einen gewissen Effekt zuschrieb, war das Echo gering und zurückhaltend. Eine Reaktion war: „Wenn uns die Grundlagenforschung sehr starke Gründe gibt zu glauben, dass ein Medikament **nicht** effektiv ist, dann ist es angemessen, bei der Interpretation augenscheinlich positiver klinischer Ergebnisse sehr vorsichtig zu sein“ (52)⁸. Diese Aussage ist absolut korrekt, wie das Beispiel des Rasierens zeigt. Man sollte generell sehr vorsichtig sein. Dass diese positive Meta-Analyse der Homöopathie keine wesentliche Sympathie eingebracht hat liegt an der grundlegenden Einstellung der wissenschaftlichen Zeitschriften. Als etwa 10 Jahre später der *Lancet* eine

zweite, negative Meta-Analyse zur Homöopathie veröffentlichte (54), begann eine regelrechte Hetzjagd. Der *Lancet* erhob gar die Forderung auf ein „Ende der Homöopathie“ (34).

Bevor auf die Studien genauer eingegangen wird, zunächst eine andere Begebenheit. Kurz nachdem der *Lancet* das Ende der Homöopathie herbei beschworen hatte, veröffentlichte das *British Medical Journal* (das ebenfalls zu den renommiertesten medizinischen Zeitschriften gehört, deren Veröffentlichungen also allgemein zur Kenntnis genommen werden) eine Meta-Analyse über die Wirksamkeit von Anxiolytika. Diese Meta-Analyse zeigte, dass Anxiolytika nicht besser wirken als Placebos (42). Auch dies war eine Neuauflage einer früheren Meta-Analyse, wobei schon 7 Jahre zuvor keine bessere Wirkung als durch Placebos festgestellt wurde (32). Nun kam aber niemand auf die Idee, das Ende der anxiolytischen Therapie zu verkünden. Parteilichkeit ist nicht weiter schlimm. Es ist in Ordnung, die eigene Meinung parteiisch vorzutragen. Wir sollten es nur nicht als Wissenschaft verkaufen.

Oft ist es ja so, dass die vehementesten Kritiker der Homöopathie keine praktisch tätigen Ärzte sind. Manche sind Physiker, Biologen, Statistiker oder Journalisten, die nur eine geringe oder keine Ahnung von den praktischen Herausforderungen der Allgemeinmedizin und auch kein Verständnis des Spannungsverhältnisses Individuum – Kollektiv haben. Sie verlassen sich auf simple Statistiken ohne Kenntnis deren Begrenztheit.

Zwei Meta-Analysen

Die erste Meta-Analyse zur Homöopathie von Linde et al. war eine Übersicht über eine Vielzahl von Studien verschiedenster Qualität, teilweise auch reine Beobachtungsstudien. Dies ist nicht unproblematisch, da diese, entsprechend den strengen Kriterien der EBM, von eher geringer Aussagekraft sind. Die zweite Meta-Analyse von Shang et al. (auch Egger-Studie) konzentrierte sich auf größere, randomisierte und doppelverblindete Studien. Das ist statistisch aussagekräftiger, aber, wie bereits gesagt, für individuelle Verläufe weniger relevant, was bei regulativen Therapien noch viel mehr zum Tragen kommt. Man könnte sa-

gen, dass die beiden Meta-Analysen an den verschiedenen Enden des Spektrums individuelle Relevanz versus statistische Robustheit anzusiedeln sind. Aber bei der Ausschlichtung der Egger-Studie wurde auf eine solch differenzierte Betrachtungsweise verzichtet.

All dies bliebe im Rahmen des normalen wissenschaftlichen Diskurses, wo Standpunkte beschrieben, Analysen erstellt und Schlussfolgerungen gezogen werden.

Eine genaue Analyse der Egger-Analyse zeigt jedoch, dass sie wissenschaftlich unzureichend ist. Sie weist eine ganze Reihe von Fehlern und Unregelmäßigkeiten auf. So wurden beispielsweise 110 Studien für diese Meta-Analyse herangezogen. Davon wurden 102 wegen wissenschaftlicher Mängel verworfen. Die letzte Auswertung basierte somit auf 8 Studien. Die Autoren verraten aber nicht, welche Studien sie verworfen haben und auf welche sie sich stützen. Das ist ein unerhörtes Vorgehen. Erst durch die internationale Kritik veröffentlichten die Autoren vorübergehend diese Daten auf der Internetseite ihres Instituts. Dabei zeigte es sich, dass sie beispielsweise Studien herangezogen hatten, die mit der Homöopathie gar nichts zu tun haben. Aber das ist nur eine der Ungeheimheiten dieser Meta-Analyse (detailliert bei 16).⁹ Eine mit solchen Mängeln behaftete Studie wäre niemals veröffentlicht worden, wenn sie ein positives Ergebnis für die Homöopathie gezeigt hätte. Oder sie wäre vom Herausgeber nachträglich zurückgezogen worden.

Unter einem nachträglichen Zurückziehen versteht man einen Vorgang, bei dem sich die Herausgeber einer Zeitschrift von der eigenen Publikation distanzieren und sie als nicht veröffentlicht betrachten. Dies geschieht, wenn nachträglich zu viele Zweifel an der Richtigkeit des Artikels bestehen oder er gefälscht wurde. Ein solcher Fall waren die Vorgänge um das „Wassergedächtnis“. *Nature* veröffentlichte einen Artikel, in dem die Arbeitsgruppe um den Wissenschaftler Benveniste bewiesen zu haben glaubte, dass der Prozess der homöopathischen Potenzierung eine Veränderung der Wasserstruktur bewirke (13). Der Artikel führte zu einiger Erregung in der wissenschaftlichen Gemeinschaft, wurde aber von *Nature* zurückgezogen, als andere Arbeitsgruppen die Ergebnisse von

Benveniste nicht reproduzieren konnten und eine Inspektion seines Labors gewisse Unregelmäßigkeiten gezeigt hatte (29). Eine vergleichbare Reaktion wäre für die Egger-Meta-Analyse durchaus angebracht gewesen. Aber es gibt in den wissenschaftlichen Zeitschriften eine erhebliche Parteilichkeit gegen die Homöopathie, wobei bevorzugt negativ berichtet wird. Dies mag zwar nicht den realen Zustand der derzeit verfügbaren Evidenz darstellen, hat aber erheblichen Einfluss auf die Wahrnehmung der Homöopathie in der Wissenschaft und Politik.

Georg Ivanovas

Zitierbar: Ivanovas G (2009): Kritik der reinen Evidenz. Homöopathie in der evidenzbasierten Medizin – Teil 1, Homöopathie KONKRET 3.9: S. 10-18. Die Literaturliste und weiterführende Links finden sich auf der Wissenschaftsseite der Homöopathie KONKRET unter: <http://www.homoeopathie-konkret.de/wissenschaft.html>

¹ Claude Bernard, einer der Väter der modernen Medizin, spricht von der alltäglichen Illusion über die Wirksamkeit unserer Therapien, wobei er den Aderlass als Beispiel nimmt, dessen Unwirksamkeit durch Béchard nachgewiesen worden war (frz: S. 342, engl: S. 195). In der klassischen Geschichtsschreibung der Medizin ist der einzige Verdienst Hahnemanns übrigens, dass er schon 70 Jahre vor Bernard und Béchard heftig gegen den Aderlass und andere schwächende Praktiken gekämpft hat. Nicht erwähnt wird in der Regel, dass Hahnemann schon vor Pinel den „Geisteskranken“ ihre Ketten abgenommen hat. Dass dies vergessen wird mag daran liegen, dass er mit zwei Fällen nicht gerade umfangreiche Studien auf diesem

Gebiet durchgeführt hat. Seine Heilungsrate von 50% ist deshalb nur wenig aussagekräftig. Aber allein dieses Vorgehen ist eine Fußnote in der Geschichte der Medizin wert und charakterisiert Hahnemann als einen klassischen Vertreter der so genannten „romantischen Medizin“, einer Medizin, bei der der Mensch in seiner Gesamtheit und Würde im Vordergrund stand.

² Originalzitat aus der französischen Ausgabe von 1865 : «Les coïncidences, dit-on, peuvent jouer dans les causes d'erreurs dans la statistique un si grand rôle, qu'il ne faut conclure que d'après des grands nombres. Mais la médecine n'a que faire de ce qu'on appelle *la loi des grands nombres*, loi qui, suivant l'expression d'un grand mathématicien, est toujours vraie en général et fautive en particulier.» (7) S. 242-243

Zitat aus der englischen Übersetzung: "It is said that coincidence may play so large a part in causes of statistical errors, that we should base conclusions only on large numbers. But physicians have nothing to do with what is called the law of large numbers, a law which, according to a great mathematician's expression, is always true in general and false in particular." (7) S. 138

³ "The development of oral contraceptives stands as a major advance in women's health in the past century."

⁴ Ghostwriting scheint eine ganz übliche Praxis in der Medizin zu sein. Der Umfang und die Bedeutung dieses Verhaltens wird derzeit von verschiedenen Seiten genauer unter die Lupe genommen (Götzsche et al. 2009; Wislar et al. 2009).

⁵ "Extension of the marketing arm of pharmaceutical companies."

⁶ "Lectures are given on the same topic by the same speakers all over the world for years, and the free debate and exchange of contrasting opinions is, de facto, banned. It was not the same 20 years ago, but in the past 15 years this trend has become a rule, to the damage of medicine and freedom."

⁷ Bei dem Ausdruck „evidence biased medicine“ handelt es sich um ein angelsächsisches Wortspiel, bei dem das Wort „based“ (fundiert, basierend auf) durch das Wort „biased“ (tendenziell, parteiisch) ersetzt wurde.

⁸ "If the basic sciences gave us very strong reason to believe that a drug would **not** be effective, then it is appropriate to be very cautious when interpreting apparently positive clinical results."

⁹ Siehe Linksammlung auf der Homepage der **Homöopathie KONKRET**.