

# Evidenz-basierte Medizin Evidenz ohne Basis

**E**videnz-basierte Medizin gilt sozusagen als Goldstandard der Medizin. Sie wird der klinischen Praxis jedoch nicht gerecht. Hier sollen einige Gründe aufgezeigt werden, weshalb der statistischen Medizin ein eher untergeordneter Rang zusteht.

## EBM und Placebo

Wird in einer Studie festgestellt, dass ein Mittel nicht wirksamer ist als ein Placebo, so heißt das nicht, dass das Mittel unwirksam ist, sondern nur, dass es unter den Untersuchungsbedingungen nicht wirksam ist oder nicht sehr oft oder nicht in der Weise, wie gemessen wird. Solche Studien haben dann meist zur Folge, dass solche Mittel nicht mehr verordnet werden dürfen oder privat bezahlt werden müssen. Damit ist aber nichts gewon-

nen. Bei milden Therapeutika, wie vielen Phytopharmaka, ist eine große Differenz zum Placebo von Anfang an nicht zu erwarten. Es werden stattdessen „wirkungsvollere“, d.h. nebenwirkungsreichere und teurere Medikamente verordnet. Da aber auch bei vielen „wirksamen“ Medikamenten der Unterschied zum Placebo nicht allzu groß ist, wie beispielsweise bei den Antidepressiva (10, 14), oder durch Tricks (17) und obskure Prozentangaben geschönt wird (8, 9), könnte man EBM als eine inquisitorische Methode betrachten, an welches Placebo der niedergelassene Arzt glauben darf. Fazit: EBM ist in weiten Bereichen weder für den Patienten hilfreich, noch ist es kostengünstiger.

Da nützt es nichts, über die Zustände zu klagen oder eine rationale Medizin (11) einzufordern, denn auch die Wissenschaftlichkeit von EBM ist in vieler Hinsicht fraglich.



Georg Ivanovas  
Milatos  
GR-72400 Neapoli  
Griechenland/Kreta

### Die Wissenschaft der Wahrscheinlichkeiten

Radikale Kritiker (3) stellen der Wissenschaftlichkeit medizinischer Forschung ein Armutszeugnis aus: „Forschung unter (Fünf-Prozent-) Niveau“. Aus Gründen der Wahrscheinlichkeitsrechnung ließe sich die Irrtumswahrscheinlichkeit klinischer Studien gar nicht berechnen und sei zudem recht hoch. Statistik sei eine „Sackgasse für die Forschung“. Diese Art von Zweifel, auch von anderen Statistikern geäußert (8), ist für Praktiker nur schwer verständlich. Aber selbst wenn man annimmt, dass Statistiken in ihren Aussagen halbwegs zuverlässig sind, gibt es noch genug andere Einwände.

### Von der Willfährigkeit der Statistiken

Ganz abgesehen von offensichtlichen Fälschungen (5, 9) sind es die Abhängigkeit der Studien von der Industrie und das Bedürfnis der Forscher, sich zu profilieren, die Anlass zur Kritik geben (15). Selten werden solche Extreme bekannt, dass unerwünschte Ergebnisse von der Industrie unterdrückt werden sollen (1). Die Spekulation darüber, wie groß der Eisberg unter solchen Spitzen ist, bleibt jedem selbst überlassen. Das Thema der vielfältigen Einflussnahme der Pharmaindustrie war erst kürzlich das Schwerpunktthema einer Ausgabe des *British Medical Journal* (4).

Zu glauben, dass medizinische Forschung dem Erwerb von wahrer Erkenntnis verpflichtet ist, ist genauso blauäugig wie zu glauben, dass Sexualität ein Ausdruck von wahrer Liebe sei. Statistische Medizin ist begrenzt durch die Redlichkeit der Forscher. Da Statistiken besonders anfällig für Manipulationen sind, können sie allenfalls mit Vorbehalt akzeptiert werden. Aber auch formale Gründe setzen statistischer Forschung enge Grenzen.

### Kleine Studien – großes Leid

Kleine Studien haben das Problem der begrenzten Fallzahl und der kurzen Dauer. So konnte in einer (kleinen) Studie über die Therapie der Epicondylitis (22) gezeigt werden, dass eine Injektion mit Kortison kurzfristig eine deutliche Verbesserung des Krankheitsbildes mit sich bringt. Langfristig ist sie aber der Physiotherapie oder dem reinen Abwarten erheblich unterlegen. Durch das Spiel mit dem Zeitrahmen lässt sich also beweisen, was immer man will. Das zeigen die langfristigen Untersuchungen, die in letzter Zeit vermehrt publiziert werden. Hochkalorische Ernährung von Frühgeborenen (große Studie) führt nach 13-16 Jahren zu erhöhter Insulinresistenz, ist also langfristig gesundheitlich ungünstiger (21). Gestillte Kinder weisen höhere TC- und LDL-Werte auf (große Studie), im Erwachsenenalter sind ihre Werte jedoch niedriger als bei Nichtgestillten (16). Allein die Dauer solcher Untersuchungen sprengt den zeitlichen Rahmen fast aller Studien, auf die wir unser „Wissen“ gründen. Wenn das Prinzip „hilft kurzfristig – schadet langfristig“ weiter verbreitet wäre, dann wären kurze Studien sozusagen eine Anleitung zur Schädigung.

Dazu kommt manchmal ein Unverständnis des Prozesses. Als Ende der 80er NaF die Standardtherapie bei Osteoporose war, kam fast jede Woche eine neue Studie auf unsere Schreibtische, die die überlegene Wirkung dieser Therapie bewies. In einer großen Studie (18) wurde später nachgewiesen, dass der Knochen unter NaF zwar dichter, aber brüchiger wird. Der Surrogatparameter „Knochendichte“ war eben nur sehr eingeschränkt zur Diagnosestellung der Osteoporose tauglich.

Dies ist kein Einzelfall. Großstudien bringen regelmäßig solche Überraschungen, die zur Abkehr von Verschreibungsgepflogenheiten führen. Häufig eingesetzte Mittel werden

dann wegen einschneidender Nebenwirkungen in ihrer Indikation eingeschränkt. Die CAST-Studie (7) wies vermehrt Todesfälle durch Antiarrhythmika nach, die WHI-Studie dasselbe durch die übliche Hormontherapie, wie sie jahrzehntelang evidenzgeprüft empfohlen wurde. Manchmal ist es erst die langjährige Praxis, die, wie bei Baycol/Lipobay, schwerwiegende Nebenwirkungen zeigt. Da solche Komplikationen in der Praxis aber nur dann bemerkt werden, wenn sie häufig auftreten, muss man von einer Vielzahl seltener, schwerwiegender Nebenwirkungen ausgehen, die niemals mit einer bestimmten Medikation in Verbindung gebracht werden, weil diese Art der Komplikation für dieses Mittel bisher nicht beschrieben ist und/oder weil sie erst nach langer Arzneimittelleinnahme auftritt.

Dieses strukturelle Problem jeder Arzneimitteltherapie lässt sich im Prinzip nur durch mehrere, große, lang dauernde Studien einigermaßen in den Griff bekommen. Allein dieses Vorgehen dürfte sich EBM nennen. Jede andere Form der Evidenz führt zu keiner sichereren und gesünderen Arzneimittelverordnung.

Andererseits haben Großstudien das Problem, dass aufgrund ihrer Inhomogenität oft nicht klar ist, ob die gefundenen Aussagen auch auf den aktuellen Patienten angewandt werden können. „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.“ (3)

### Willkürliche Evidenz

Außerdem lösen Großstudien nicht das logische Problem der kausalen Beziehung. Statistiken erzeugen Werte und Korrelationen, aber keine

Kausalitäten. Welche Art der Relation einer Korrelation zu Grunde liegt, darüber kann eine Statistik nichts aussagen. Das beliebteste Beispiel in der statistischen Literatur ist die Korrelation der Abnahme der Storchenpopulation, die signifikant mit dem Geburtenrückgang korreliert. Eine noch bessere Korrelation zeigt die Abnahme des Kartoffelkonsums mit der Zunahme des Stromverbrauches (19). Obwohl dieser Zusammenhang statistisch unbestreitbar besteht, so muss man doch recht komplexe Erklärungsmuster bemühen, diese beiden Werte in Beziehung zu setzen. Was machen wir aber nun aus der Caerphilly-Studie (6), wo bei 2.438 Männern, die über 20 Jahre beobachtet wurden, bei jenen Männern, die sich seltener als einmal täglich rasierten, eine um 70 % erhöhte Inzidenz an Schlaganfällen und eine um 30 %

erhöhte Mortalität festgestellt wurde. Nach üblicher Forschungslogik müssen wir eine kausale Beziehung annehmen und Konsequenzen ziehen. Zunächst müssen Rasierapparate per Kassenrezept verordnet werden. Dann muss es Aufklärungskampagnen über den Wert des Rasierens geben. Man müsste sich überlegen – und in kleinen Studien prüfen –, ob Enthaarungscremes das Problem vielleicht lösen, denn kein Bart – kein Rasierproblem!

Was sich zunächst wie eine Satire anhört, stellt sich schnell als die Real satire der medizinischen Forschung heraus.

Viele therapeutische Strategien, die aus statistischer „Erkenntnis“ abgeleitet werden, bewegen sich genau auf diesem logischen Niveau. Da es sich dabei jedoch meist um Diabetes oder rheumatoide Arthritis handelt, fällt es nicht so auf.

Das wissenschaftstheoretische Problem liegt darin, dass die Bewertung der Korrelationen nach einem gewissen Gutdünken erfolgt. Es gibt schlicht kein anderes Verfahren. Statistiken erzeugen niemals „wahre“ Aussagen und statistische Evidenz ist in gewisser Weise immer eine „Evidenz aus Offensichtlichkeit“. Statistiken können nur das ergeben, was eine von außen kommende Logik zulässt. Thesen können sich als wahrscheinlicher oder unwahrscheinlicher zeigen. Zu glauben, dass Statistiken neues Wissen erzeugen, ist wie das „Suchen von Ostereiern, die man selbst versteckt hat“ (20).

Doch damit nicht genug. Um über die Wirksamkeit von Therapien zu urteilen, die nach statistischer Erkenntnis erdacht wurden, vergehen weitere zehn Jahre. Bis dorthin sind diese Therapien durch neuere Medika-

# Titelthema

mente ersetzt, die sich, wie moderne Antipsychotika, im Nachhinein (Meta-Analyse) als auch nicht wirksamer herausstellen als die alten (13). Zumindest nicht nach EBM-Kriterien.

## Wertlose Messung

Hinter der willkürlichen Bedeutungskonstruktion einer Statistik verbirgt sich ein zentrales Problem der Wissenschaftstheorie. Eine Information hat keine Bedeutung aus sich selbst heraus. Ihren Wert erhält sie erst durch den Kontext (2). Dieses seltsame Phänomen spielt in der Medizin eine große Rolle. So sagt ein positiver HIV-Test zunächst nicht sehr viel aus, vor allem nichts darüber, ob der Patient an AIDS erkrankt ist oder nicht (8). Es ist dies eine dieser statistischen Kniffeleien: Die Wahrscheinlichkeit, ob ein HIV-Positiver wirklich erkrankt ist, hängt davon ab, welchen „Umgang“ er hat. Die einzelne Messung ist per se nichts sagend. Erst durch die Evaluierung der individuellen Charakteristika des Patienten bekommt sie ihren Wert.

Auch der Wert des Mammografie-Screenings ist umstritten. Gerade Statistiker sind eher zurückhaltend, was

deren Nutzen betrifft (8). Ob ein flächendeckender Einsatz in Deutschland wirklich zu einer Senkung der Mortalität führt (12), wird sich erst in Jahrzehnten herausstellen. Natürlich hängt die Antwort sehr stark davon ab, ob man die Mortalität an Brustkrebs untersucht, die Mortalität an diagnostiziertem Brustkrebs oder gar die Mortalität allgemein. Man könnte stattdessen – bei begrenzten finanziellen Ressourcen – dieses Geld zur Adipositas-Prophylaxe einsetzen, deren Wert gesichert ist. Wie würde dann die Statistik der Gesamtmortalität aussehen? Statistik ist sozusagen eine große Spielwiese für die verschiedensten Interessen. Statistisch sinnvolle Bewertungen hinken den Gegebenheiten immer Jahrzehnte hinterher.

## Magie der Grenzwerte

Aus Messwerten wird nach bestimmten Berechnungen, die zukünftigen Generationen vielleicht einmal als Zahlenmystik erscheinen mögen, heute aber Standardabweichung genannt werden, ein Grenzwert festgelegt. Aus der mathematischen Darstellung einer Verteilung, einer beschreibenden Gesetzmäßigkeit, werden

Interventionspunkte geschaffen. Diese stecken, unabhängig vom Einzelfall, das Territorium von Krankheit und Gesundheit ab. Es sind vorschreibende Gesetze geworden. Wie bei der Übertretung der Höchstgeschwindigkeit ein Bußgeld oder ein Strafverfahren fällig wird, so zieht ein Über- oder Unterschreiten dieser Grenzwerte eine therapeutische Intervention nach sich. Es sei denn, Polizist oder Doktor drücken mal ein Auge zu. Aber im Gegensatz zur Polizei darf der Doktor auch schon einschreiten, wenn jemand schon in die Nähe der Grenze kommt („Ihre Werte sind ziemlich hoch!“ – das heißt normal), sozusagen eine prophylaktische Strafe.

Diese „Magie der Grenzwerte“ ist nicht rationaler als die Medizin vergangener Jahrhunderte.

## Vom Kollektiv zum Individuum

Natürlich gehen die meisten Therapeuten flexibel mit solchen Grenzwerten um und betrachten den Einzelfall. Das ist weitgehend ein Ausdruck von Erfahrung. Erfahrung ermöglicht es, eine Vielzahl von Komponenten zu einem sinnvollen Ganzen zu ver-

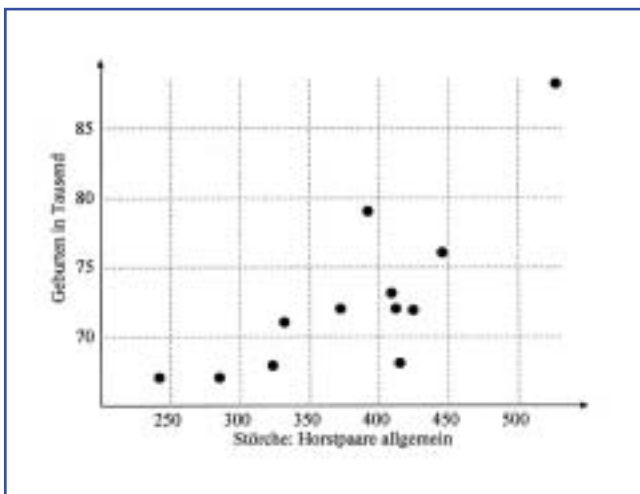


Abb. 1: Die hier dargestellte Korrelation wird zu einer Relation, wenn man dem Modell folgt, dass Störche die kleinen Kinder bringen. [aus Schwarze: 17]

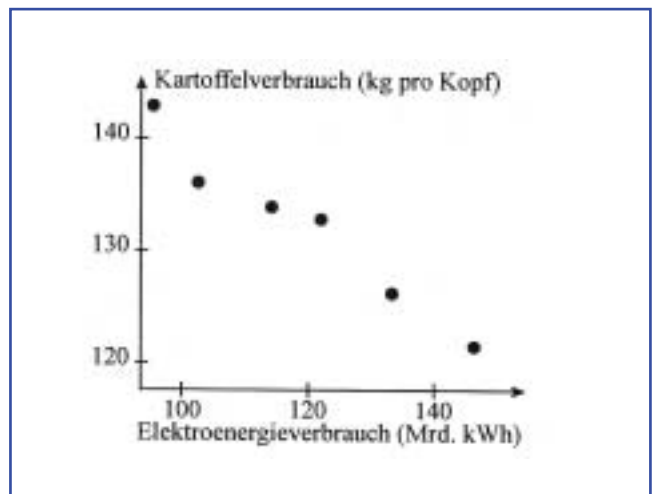


Abb. 2: Die Korrelation ist so eindeutig, dass es sich um keinen Zufall handeln kann, was jedoch keine Kausalität impliziert. [aus Schwarze: 18]

knüpfen und über statistische und Lehrbuchangaben hinaus zu individuellen Schlussfolgerungen zu kommen. Durch ein besseres Verständnis individueller Ausprägungen könnten Therapieverläufe besser interpretiert werden. Wenn eine Therapie, wie meistens, nur einem geringen Teil der Patienten nützt, zum Beispiel nur jedem Zehnten, so könnten wir bei einem besseren Verständnis individueller Reaktionen eventuell herausfinden, wem es nützt oder wem es eventuell schadet. Ein solches Vorgehen bleibt derzeit weitgehend der „ärztlichen Intuition“ überlassen. Das mag in gewisser Weise notwendig sein, ist aber wissenschaftlich nicht befriedigend. Denn „manche bezeichnen die Fehler, die sie seit 30 Jahren machen, als Erfahrung“. Was fehlt, sind nachvollziehbare und nachprüfbar Kriterien, die eine individuelle Beurteilung erlauben. Dabei gibt es in der Systemtheorie dafür durchaus ein wissenschaftliches Werkzeug. Es hat in der Psychotherapie bereits wertvolle Dienste geleistet. Auch in manchen naturheilkundlichen Verfahren wie der Kurmedizin gibt es durchaus Ansätze, aus individuellen Reaktionen ein größeres Reaktionsmuster zu erschließen.

Natürlich könnte die Wirksamkeit eines solchen Vorgehens nur statistisch erfolgen. Ob das aber aussagekräftig wäre?

## Literatur

1. Baird P, Downie, Thompson J: Clinical Trials and Industry. Science 2002; 297: 2211
2. Bateson G: Ökologie des Geistes. Frankfurt: Suhrkamp, 1990
3. Beck-Bornholdt HP, Dubben HH: Der Schein der Weisen. Reinbek bei Hamburg: Rowohlt 2003
4. BJM 2003; 326 (31. Mai)
5. DiTrochio F: Der große Schwindel. Frankfurt, New York: Campus 1995
6. Ebrahim S, Smith G, May M, Yarnell J: Shaving, coronary heart disease, and stroke. Am J Epidemiol 2003; 157: 234-238.
7. Echt DS, Liebson PR, Mitchell LB et al.: Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. NEJM 1991; 324: 781-788
8. Gigerenzer G: Das Einmaleins der Skepsis – Über den richtigen Umgang mit Zahlen und Risiken. Berlin: Berlin
9. Haffner P, Luczak H: Und ewig lockt der Ruhm. GEO 2003; 3: 120-138
10. Kirsch I, Sapirstein G. (1998): Listening to Prozac but Hearing Placebo: A Meta-Analysis of Antidepressant Medication. Prevention & Treatment 1998; Vol 1. Article 0002a
11. Koch K: Therapieentscheidung wider besseres Wissen: Warum Ärzte gegen Viren mit Antibiotika vorgehen. Deutsches Ärzteblatt 1999; 96: A-1822 (Heft 27)
12. Köhler A, Gibis B, Mühlich A: Mammografie-Screening: Flächendeckendes Angebot bereits im Jahr 2005. Dtsch Arztebl 2003; 100: A 1240-1243 (Heft 19)
13. Leucht S, Wahlbeck K, Hamann J, Kissling W: New generation antipsychotics versus low-potency conventional antipsychotics: a systematic review and meta-analysis. Lancet 2003; 361: 1581-89
14. Leuchter A, Cook I, Witte E, Morgan M, Abrams M: Changes in Brain Function of Depressed Subjects During Treatment With Placebo. Am J Psychiatry 2002; 159: 122-129
15. Mertens S: Medizinische Publikationen: Viel heiße Luft. Dtsch Arztebl 2001; 98: A-3021 (Heft 46)
16. Owen CG, Whincup PH, Odoki K, Gilg JA, PhD, Cook DG, PhD: Infant feeding and blood cholesterol. Pediatrics 2002; 110: 597-608
17. Pirisi A: Antidepressant drug trials exclude most „real“ patients. Lancet 2002; 9308: 767
18. Riggs BL, Hodgson SF, O’Fallon WM: Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. N Engl J Med 1990; 322: 802-809
19. Schwarze J: Grundlagen der Statistik. Herne, Berlin: Neue Wirtschafts-Briefe 2001, S. 18
20. Simon FB: Unterschiede, die Unterschiede machen. Frankfurt: Suhrkamp 1993, S. 31
21. Singhai A, Fewtrell M, Cole TJ, Lucas A: Low nutrient intake and early growth for later insulin resistance in adolescents born term. Lancet 2003; 361: 1089-1097
22. Smidt N, van der Windt DAWM, Assendelft WJJ, Devillé WLJM, Korthals-de Bos IBC, Bouter LM: Corticosteroid injections, physiotherapy, or a wait-and-see policy for lateral epicondylitis: a randomised controlled trial. Lancet 2002; 359: 657-662