

III. Empirische Dokumentation von Wirkung und Nutzen in der Hierarchie der Evidenz/ Empirical Documentation of Efficacy and Benefit in the Hierarchy of Evidence

Kommentar des Herausgebers

Die Medizin als Wissenschaft hat heute die Aufgabe, sich auf eine möglichst intelligente *empirische* Dokumentation von Therapie-wirkung und Therapienutzen zu verständigen. Fragen der Wirkungsmechanismen bleiben da, aber sie gehen in den Hinterkopf.

Der oft noch nicht recht wahrgenommene markante historische Wandel, welcher die neue Schulmedizin der «Evidence-Based Medicine» mit sich bringt, besteht im neuen Evidenz-Begriff. Dieser stellt Wirkungstheorien und das Verständnis physiologischer Wirkungsmechanismen auf die Stufe blosser Hypothesen und setzt die Ergebnisse kontrollierter, empirischer Studien an die Spitze der Hierarchie der Evidenz. Mit der neuen Grundlagenwissenschaft der klinischen Epidemiologie misst sich die Wissenschaftlichkeit von Medizin heute an der Qualität der empirischen Wirkungs- und Nutzensdokumentation (und nicht mehr an der pathophysiologischen Plausibilität oder Konformität einer Wirkungstheorie).

Für die Qualität empirischer Wirkungs- und Nutzensdokumentation entscheidend ist die Hierarchie der Evidenz, welche den Studien mit den geringsten Fehlermöglichkeiten am meisten Aussagekraft einräumt. Wenn zu einer gleichen Frage die Ergebnisse von *randomisiert kontrollierten Studien* vorhanden sind, dann sind diese Ergebnisse zuverlässiger und damit «wissenschaftlicher» als (möglicherweise widersprechende) Beobachtungen aus Fallstudien oder Kohortenstudien, welche systematischen Fehlern («*selection*» oder «*confounding bias*») unterliegen. Dass in der Praxis aufgrund der einseitigen Vorherrschaft pharmakologischer Forschungsfragen nicht bereits die «beste Evidenz» oder die Evidenz aus randomisiert kontrollierten Studien mit der besten Therapie verwechselt werden darf [1], stellt diese Hierarchie der Evidenz nicht im Grundsatz in Frage. Darüber hinaus bestehen bekannte Beobachtungsfehler (Gefälligkeits-Bias, Beobachter-Voreingenommenheit etc.), welche unabhängig vom Studientyp und seinem Platz in der Hierarchie ebenfalls durch geeignete Massnahmen kontrolliert werden müssen [2]. Hier spielt etwa das Verblinden der Erfolgsbeurteilung meist eine weit wichtigere Rolle als das Verblinden von Behandler und Patient (in der Doppelblindstudie). «Erfahrung»

Editorial Commentary

The mission of medical science today is to find a consensus on the most intelligent and best *empirical* way of documenting therapeutic effectiveness and therapeutic benefit. Questions about mechanisms of action will remain, but will retreat to the back of the mind.

The often not yet truly perceived, but marked historic transition associated with the new school of evidence-based medicine involves the new understanding of what constitutes evidence. This places action theories and the understanding of physiological mechanisms on the level of mere hypotheses and places the results of controlled empirical studies at the top of the hierarchy of evidence. With the new basic science of clinical epidemiology, the scientific soundness of today's medicine is measured by the quality of the empirical documentation of effectiveness and benefit (and no longer by the pathophysiological plausibility or conformity with a theory of action).

What is crucial for the quality of the empirical documentation of efficacy and benefit is the hierarchy of evidence which grants studies with the lowest possibility of bias the greatest scientific soundness. When the results of randomised controlled trials are available for a particular question, then these results are more reliable than others and thus more 'scientific' than (possibly contradictory) observations from case studies or cohort studies which are subject to systematic errors (selection or confounding bias). The fact that, in practice, the 'best evidence' or evidence from randomised controlled trials must not be confused with the best therapy [1] (e.g. due to the one-sided predominance of pharmacological research questions) does not question the basic principles of this hierarchy of evidence per se. Moreover, there are well-known potential biases (recall bias, observer bias etc.) which must also be kept in check by suitable measures, independent of study design and rank in the hierarchy [2]. This is where blinding of the outcome assessment usually plays a much more important role than blinding of therapist and patient (in double-blind studies). 'Experience' in terms of good observation can provide important scientific information especially when controlled trials or the possibility of conducting

im Sinne guter Beobachtung kann vor allem bei fehlenden kontrollierten Studien oder fehlender Möglichkeit, kontrollierte Studien durchzuführen, eine wichtige wissenschaftliche Information ausmachen. Sehr oft sind «gute Erfahrungen» aber unkritische und voreingenommene Berichte, welche den einfachsten Evidenzregeln nicht standhalten.

Mit den folgenden Beiträgen findet nicht eine vollständige und systematische Auseinandersetzung mit den Kriterien guter, aussagekräftiger klinischer Studien statt; dies ist an anderer Stelle bereits mehrfach geleistet worden [3–6] und wird auch in Kapitel V weiter ausgeführt. Der folgende Beitrag von *Gerd Büschel* et al. zeigt am Beispiel unkonventioneller Krebstherapien aber sehr schön auf, wie die Fehler der «alten» Schulmedizin auch unkonventionelle Therapieschulen durchziehen. So wie auch die «alte» Schulmedizin etwa die Mastektomie und die Früherkennungslehre aufgrund unzureichender Beobachtungen eingeführt hat [7, 8], ohne über Jahrzehnte die Krebsmortalität verbessern zu können, wird offenbar auch der «Erfolg» unkonventioneller Medikamente und der Krebs-Mehrschritt-Therapie durch Selektion der Fälle und unsorgfältige Beobachtung herbeigeredet. Auch wenn dies nicht unbedingt schlechter ist als das, was die «alte» Schulmedizin oft geleistet hat, so wird und muss aber auch diejenige unkonventionelle Medizin untergehen, die ihre Forschungsqualität mit Hilfe der Instrumente der klinischen Epidemiologie nicht verbessert. Eine Bestfall-Analyse kann die behauptete Tumorwirksamkeit dieser Therapien relativ sicher widerlegen, soweit man den «Surrogat-Marker» der konventionellen Tumorregressions-Kriterien als gültig betrachten kann. Sicher müssen bei Krebserkrankungen statt der Tumorregression allein auch wirtsspezifische Wirkungen [7] oder weitergefasste Lebensqualitäts-Endpunkte berücksichtigt werden. Es liegt aber an den Therapieanbietern, hier entsprechende Studien vorzulegen, bevor sie mit falschen Wirksamkeitsbehauptungen ihre Verfahren anbieten. Eine gute Therapie bei Krebs ist, die Ohnmacht ertragen zu lernen und zu trösten statt unnütze Behandlungen anzubieten. Diese wichtigste Herausforderung wird heute von Vertretern unkonventioneller Verfahren oft ebensowenig gemeistert wie von der «Schulmedizin».

Der Beitrag von *Markus Fähr* vermittelt einen Einblick in die Diskussion um die Psychotherapieforschung im deutschsprachigen Raum. Allerdings bringt er wenig Licht in die Frage der möglichst validen Gestaltung empirischer Forschung. *Fähr* lässt uns im unklaren, weshalb Grawe et al. mit ihrer Kritik an der mangelnden Dokumentation des Nutzens von Langzeitpsychotherapie falsch liegen sollen, wir müssten es ihm mangels methodisch differenzierter Erörterung einfach glauben. Ein wichtiges Prinzip von Wissenschaft ist die «externe Validität», und gute Therapien können im falschen Moment oder beim falschen Patienten schaden, insbesondere wenn Studienergebnisse auf Gruppen mit unterschiedlicher Krankheitsausprägung übertragen werden [7]. Es geht deshalb darum, den Stellenwert und die Indikationen von Psychotherapie (in Konkurrenz zu anderen Therapien) genauer zu bestimmen, um zu wissen, wo genau sie Nutzen hat (nicht nur dass sie Nutzen hat) und wo sie auch schadet. Die von *Fähr* in den Vordergrund gestellt-

controlled studies are lacking. Very often, however, ‘good experience’ amounts only to uncritical and biased reports which do not stand up to the simplest rules of evidence.

The following contributions do not provide a complete, systematic presentation of the criteria of sound clinical studies; this has been done repeatedly elsewhere [3–6] and will also be outlined in chapter V. The following report by *Gerd Büschel* et al. uses the example of unconventional cancer therapies to very aptly illustrate how the mistakes of the ‘old classical medicine’ have also infiltrated unconventional schools of medicine. Just as ‘old classical medicine’ has introduced mastectomy and the dogma of early detection based on insufficient and biased observations [7, 8], without being able to improve cancer mortality for decades, the ‘success’ of unconventional drugs and multistage cancer therapy is also obviously an artefact of patient selection and careless observation. Even if this is not necessarily worse than what ‘old classical medicine’ has often done, this type of complementary medicine will and must also disappear if it does not improve its research quality by employing the instruments of clinical epidemiology. The best-case analysis by *Büschel* et al. seems to conclusively refute claims that the therapies studied have an effect, as far as conventional tumour regression criteria as surrogate markers are validly measuring the success of treatment. Assessing cancer therapy would not only need to consider tumour regression, but also possible host-specific effects [7] and measures of quality of life. Here, however, it is the job of those propagating these therapies to present us with the proper studies before selling unproven therapies. A good ‘therapy’ for cancer patients always is to learn how to cope with helplessness and to comfort instead of prescribing useless or harmful treatments. This most important challenge, however, is often not met by conventional doctors as well as by practitioners of unconventional therapies.

Markus Fähr gives some insight into the discussions about psychotherapy evaluation (in German-speaking countries). However, his contribution sheds little light on the question of the best possible, valid design of empirical research. *Fähr* leaves us in the dark about why Grawe et al. with their study criticizing the lack of evidence for a benefit of long-term psychotherapy are supposed to be wrong. We must simply believe him as his presentation lacks methodological differentiation. An important principle of science is ‘external validity’; good therapies given at the wrong time or to the wrong patient can harm, especially when study results are applied to groups with diseases of different degrees of severity [7]. Therefore, it is important to determine the place of psychotherapy accurately (in competition with other therapies) in order to know exactly *where* it is beneficial (not only that it is beneficial) and where it may also harm. Then, we must classify ‘naturalistic follow-up studies’, used by *Fähr* as evidence for the superiority of long-term therapy, within the ranks of non-controlled studies that provide us at the most with better hypotheses, but no conclusive knowledge. Observational studies always suffer from potential selection artefacts (e.g. healthy complier bias). Patients who receive or comply with high-frequency long-term therapy certainly differ from patients who undergo a brief therapy regimen. How is one to know

ten «naturalistischen Katamnesestudien» müssen wir in die Reihe unkontrollierter Studien einordnen, die uns höchstens bessere Hypothesen, aber kein sicheres Wissen vermitteln. Bei Katamnesestudien lässt sich ein Selektionsartefakt (z. B. «healthy complier bias») als Störfaktor nie sicher vermeiden. Patienten, die eine hochfrequente Langzeittherapie mitmachen, unterscheiden sich fast sicher von Patienten, die eine Kurztherapie bekommen bzw. mitmachen. Wie soll man hier wissen, ob der Unterschied im Therapieverfahren oder einfach die Patientenselektion zu den unterschiedlichen «outcomes» führt. Ohne (mindestens exemplarische) kontrollierte Studien kann nicht einmal ausgeschlossen werden, dass ein Selektions-Bias einen möglichen Schaden der hochfrequenten Langzeittherapie verdeckt. *Fäh* unterliegt dem Missverständnis, dass in der Medizin die kontrollierte Studie als die «exakteste» Studie gilt. Sie ist «nur» die Studie, welche alle möglichen alternativen Hypothesen für einen beobachteten Unterschied zwischen Therapie und Nicht-Therapie «unter Kontrolle» hält. Es wäre interessant zu erfahren, wo eine praxisgerechte, individualisierte (Langzeit-)Psychotherapie in einer randomisiert kontrollierten Studie zu einer Leidensverminderung geführt hat, die zudem klinisch bedeutsam ist («Number needed to treat?»). «Exakt» wäre eine solche Studie keineswegs, sie könnte uns nur über den durchschnittlichen Stellenwert der Psychotherapie bei einem bestimmten Problem Auskunft geben, nicht aber über die exakte Wirkung im Einzelfall. Die Durchschnittswirkung randomisiert kontrollierter Studien ist dennoch eine unabdingbare Information, denn wenn die Durchschnittswirkung fehlt, kann auch keine Wirkung in Einzelfällen mehr angenommen werden (ausser es gibt begründete Untergruppen-Hypothesen; in diesem Fall wäre eine neue kontrollierte Studie mit der «richtigen» Untergruppe fällig). Die von *Fäh* postulierte «Intermethoden-Konvergenz» (in impliziter Ablehnung der Hierarchie der Evidenz) ist deshalb ein Trugschluss, der mindestens dann erkannt werden kann, wenn pragmatische, kontrollierte Studien als Möglichkeit ins Bewusstsein gelangen, die auch heuristisches «Messen» nicht ausschliessen und individualisierte Therapien erlauben [9]. Es ist hilfreich, zur Kenntnis zu nehmen, dass auch die konventionelle «somatische» Medizin zum grössten Teil heutigen wissenschaftlichen Erfordernissen nicht entspricht [10], und es ist deshalb unnötig, methodische Verteidigungskonstruktionen für die Psychotherapie zu erfinden, die nicht durchdacht sind, auch wenn sie auf den ersten Blick viel Richtiges enthalten. Unsere Arbeit ist nicht mehr, Schulen zu «beweisen», sondern die besonderen Situationen herauszuarbeiten, wo das eine oder andere gültig und wo es unnütz ist. Wenn analytische Langzeittherapien nicht in Studien ohne Selektions-Bias untersucht sind, dann gilt es, diese Studien nachzuholen und daraus zu lernen, und nicht, eine Verteidigungsmethodologie zu basteln, die einen Erkenntnisgewinn nicht richtig leisten kann. Die Möglichkeiten pragmatischer kontrollierter Studien sind keineswegs ausgeschöpft und werden von *Fäh* nicht einmal in Betracht gezogen. In diesem Sinn hilft dieser einseitige Beitrag, das eigentliche Ziel dieses Heftes klar herauszuarbeiten und führt auf die Beiträge im letzten Kapitel hin. Eine Erkenntnis wird sein, dass in allen Therapiebereichen unnütze Therapien aufgegeben werden müssen – in einigen mehr, in anderen weniger.

whether it is the difference in the method of treatment or simply the patient selection and thus the difference in the compared groups that leads to different outcomes? Without controlled studies (at least some exemplary ones) we cannot even rule out if a selection bias is masking a potentially harmful effect of high-frequency long-term therapy. *Fäh* is mistaken in that the controlled trial ranks as the ‘most exact’ type of study in medicine. It is ‘only’ the study design that keeps all possible alternative hypotheses for an observed difference between therapy and non-therapy ‘under control’. Therefore, we would be interested to know where a practically sensible, individualized (long-term) psychotherapy regimen has been shown in a randomised controlled study to reduce illness to an extent that is also clinically relevant (number needed to treat?). Such a study would in no way be ‘exact’, it would only give us information about the average effect of psychotherapy for a certain illness, but not about the exact effect in the individual patient. The average effect of randomised controlled studies, nevertheless, provides us with indispensable information since, when the average effect is lacking, there is also no effect in individual patients (unless well-founded subgroup hypotheses are available; in this case, a new controlled study of the ‘right’ subgroup would be needed). The ‘intermethod convergence’ postulated by *Fäh* (while implying a rejection of the hierarchy of evidence) is a fallacy that is at least recognisable once one has become aware of the possibility of conducting pragmatic, controlled studies that do not exclude even heuristic ‘measuring’ and that allow individualized therapies [9]. It worth noting that also conventional ‘somatic’ medicine is not evidence-based to a great extent [10], and it is therefore unhelpful to invent ‘defense methodologies’ (such as the ‘intermethod convergence’) which are not thought out even if they at first glance contain many valid points. The current task of science is no longer to prove therapy schools, but to document the specific situations in which this or that is valid or of no use. If analytical long-term psychotherapy is not yet evaluated in controlled studies which are free of selection bias, it is time to do this and not to propose defense methodologies which will not produce the knowledge and insight we need. The potential of pragmatic controlled trials is by no means exhausted and is not dealt with by the author. In this context, this one-sided report helps to clearly extract the essence of the actual aim of this book and points the way to the contributions in chapter V. It will be understood with time that in all areas of therapy unhelpful treatments will have to be abandoned – in some cases more, in some cases less.

Literatur/References

- 1 Feinstein AR, Horwitz RI: Problems in the 'Evidence' of 'evidence-based medicine'. *Am J Med* 1997;103:529–535.
- 2 Edwards RA: Research and the goal of improving patient care. *Forsch Komplementärmed* 1998;5(suppl 1):116–120.
- 3 Sackett DL, Richardson WS, Rosenberg W, Haynes RB: *Evidence-Based Medicine – How to Practice & Teach EBM*. New York, Churchill Livingstone, 1997.
- 4 Trisha Greenhalgh: *How to Read a Paper. The Basics of Evidence-Based Medicine*. London, BMJ Publishing Group, 1997.
- 5 Riegelman RK, Hirsch R: *Studying a Study and Testing a Test: How to Read the Medical Literature*. Boston, Little, Brown and Company, 1989.
- 6 Steurer J, Bucher HC, Schmidt JG: Evidence Based Medicine: Kritische Beurteilung einer «Therapiestudie». *Praxis* 1997;86:1603–1607.
- 7 Schmidt JG: Die Vision einer pragmatischen klinischen Forschung oder das Ende der Diskussion über «Placebo» und «spezifische Wirkungen». *Forsch Komplementärmed* 1998;5(suppl 1):102–111.
- 8 Schmidt JG: Was bringt das Screening beim Brustkrebs? *Arch Gynecol Obstet* 1996;259:S178–S198.
- 9 Holle R, Pritsch M: Randomisierte Studien mit individualisierten Therapien: Erfahrungen aus der Psychotherapieforschung. *Forsch Komplementärmed* 1996;3:78–84.
- 10 Field MJ, Lohr NN (eds): *Institute of Medicine Guidelines for Clinical Practice: From Development to Use*. Washington DC, National Academy Press, 1992.