

# Gibt es Gründe für Pluralistische Evaluationsmodelle? Limitationen der Randomisierten Klinischen Studie

Gunver S. Kienle

Institut für angewandte Erkenntnistheorie und medizinische Methodologie, Bad Krozingen

## Zusammenfassung

Der randomisierten klinischen Studie (RCT) wird heute der höchste Evidenzwert zugesprochen, andere Evidenztypen werden selten in Bewertungen integriert. Unter Idealbedingungen zwar perfekt, haben RCTs in der klinischen Realität jedoch zahlreiche Limitierungen. Ihre Priorisierung führt zu Verzerrungen in der Gesundheitsversorgung durch Kommerz-Bias, Karriere-Bias, Bias der großen Zahl, Mediokritäts-Bias, Bias durch Therapiepriorität, zu Diskrepanzen von Forschung und medizinischem Alltag, zu ethischen Problemen, außerdem zu divergenten und asymmetrischen Ergebnissen. Zudem verführt die vermeintlich einfache Beurteilung mittels RCTs zu ihrer Instrumentalisierung und zu ärztlicher Fremdbestimmung durch Fachfremde, insbesondere Bürokraten. Um den realen, komplexen Gegebenheiten der medizinischen Versorgung gerecht zu werden, sollten die ärztliche Urteilskraft professionalisiert und in die medizinische Erkenntnis re-integriert, pauschale Evidenzhierarchien aufgehoben und eine Informationssynthese aus verschiedenen (qualitätsgeprüften) Evidenztypen entwickelt werden.

**Sachwörter:** Randomisierte klinische Studie – Public Health – Ärztliche Urteilskraft – Evidence-based Medicine – Cognition-based Medicine – Bias

## Is There a Rationale for Pluralistic Evaluation Models? Limitations of Randomized Clinical Trials

### Summary

Randomized clinical trials (RCTs) are believed to deliver the highest level of clinical evidence, and other types of evidence are rarely integrated in therapy assessments. Under ideal circumstances, RCTs are perfect tools of therapy evaluation; however, in clinical reality they suffer from a variety of limitations. Their prioritization can distort public health regulation because of commercial bias, career bias, bias of large numbers, mediocrity bias, and bias of therapy preferences; also, they can lead to discrepancies between clinical research and routine medical practice, to ethical problems, and to divergent and asymmetrical results. Moreover, the seemingly simple therapy assessment based on RCT results bears the danger of their endorsement and misuse by non-medical personnel, especially bureaucrats. To meet the demands of the complex reality of medical care, the physician's judgment should be methodologically professionalized and re-integrated into the process of generating medical knowledge; fixed hierarchies of medical evidence should be de-emphasized in favor of information syntheses from different types of (quality assessed) evidence.

**Key words:** Randomized clinical trial – Public health – Clinical judgment – Evidence-based medicine – Cognition-based medicine – Bias

Der Bewertung von Therapieverfahren werden heute Evidenzhierarchien zugrundegelegt, die der randomisierten klinischen Studie (RCT) und ihren Meta-Analysen den höchsten Evidenzwert beimessen. Weitere Evidenzstufen betreffen andere Studien- und Erkenntnisformen; sie werden jedoch *in praxi* nur selten berücksichtigt. Das Konzept der RCT ist heuristisch überzeugend. Eine RCT erlaubt unter Idealbedingungen eine zuverlässige Beurteilung des relativen Vorteils einer bestimmten Therapie auf einen bestimmten Krankheitsparameter. Diese Zuverlässigkeit wird ermöglicht durch Homogenisierung, Randomisierung, Standardisierung und, falls möglich, Verblindung. Jedoch bestehen die Idealbedingungen bei weitem nicht immer. In der klinischen Realität sind RCTs oft nicht durchführbar; nicht selten produzieren sie nur wenig relevante oder auch unzuverlässige Ergebnisse. Ihre Priorisierung führt in der Gesundheitsversorgung zu Verzerrungen zugunsten gesundheitsfremder Interessen (s. Tabelle 1). Dies wird im folgenden exemplifiziert.

## Kommerz-basierte Medizin durch Evidenz-Bias

“A basic assumption in guidelines development is that clinical research follows clinical relevance. In reality, much research is driven by commercial inte-

**Tabelle 1.** Eingeschränkte Eignung von RCTs.

**RCTs nur einseitig realisierbar**

- Kommerz-Bias
- Karriere-Bias
- Bias der großen Zahl
- Ethische Probleme
- Therapiepriorität, individuelle Therapie

**Unzuverlässige Ergebnisse in RCTs**

- Divergente Ergebnisse
- Asymmetrie der Ergebnisse
- Mediokritäts-Bias

**RCTs oft nicht praxisrelevant**

- Diskrepanz von Forschung und medizinischem Alltag
- Fortschrittsbehinderung durch Formalisierung und Bürokratisierung

**Gesundheitspolitische Probleme**

- Vermeintlich einfache Beurteilung verführt zur Instrumentalisierung durch Fachfremde, zur Zentralisierung der Medizin

rests.“ [1] – RCTs sind teuer; abhängig von der untersuchten Fragestellung und der Laufzeit übersteigen die Kosten in der Regel siebenstellige Beträge und können 30 Millionen Dollar [2] und mehr erreichen.

Wegen hoher Kosten bei zugleich geringer staatlicher bzw. gemeinnütziger Förderung wandert die klinische Forschung zunehmend in die Domäne der pharmazeutischen Industrie ab, wo sie Zulassungs- und Marketingsinteressen gehorcht. Die Folge ist, dass prioritär nur Therapien erforscht werden, die patientierbar und Gewinn versprechend sind. Gemeinhin investiert die pharmazeutische Industrie nur in Medikamente, deren geschätzter Umsatz über 300 Millionen Pfund pro Jahr liegt [3]. Viele Erfolg versprechende Therapien werden deshalb nur schlecht oder nie überprüft, zumal wenn sie sich nicht durch Patente schützen lassen.

Die kostenaufwendige moderne Forschungslogistik ist realisierbar bei Erkrankungen, die große Menschenmassen betreffen (Volkskrankheiten wie die koronare Herzerkrankung, mit vielen Menschen sowohl für Studien als auch im späteren Absatzmarkt). Sie begünstigt pharmazeutische Unternehmen, die mit wenigen Medikamenten großen Umsatz machen. Verlierer sind dagegen Therapien ohne Aussicht auf breite Vermarktung bzw. ohne finanzstarke industrielle Rückendeckung: nicht-pharmakologische Therapien (z.B. Physiotherapien, Kreativtherapien, Psychotherapien, chirurgische Verfahren, manuelle Therapien, alle Anwendungs- und Hinwendungs-orientierten Verfahren, Verhaltensänderungen), Therapien tropi-

scher Erkrankungen („neglected diseases“), Behandlungen seltener Erkrankungen, Langzeitbehandlungen chronischer Erkrankungen, Antidote gegen Vergiftungen, Impfstoffe, Therapien von Kindern, komplementärmedizinische Verfahren (wo viele nicht-pharmakologische Therapien eingesetzt werden und wo aufgrund der Systemlogik die pharmazeutischen Hersteller hunderte oder tausende Arzneispezialitäten vorhalten müssen). Ungeachtet ihrer Wirksamkeit besteht für alle diese benachteiligten Therapien die Gefahr, dass sie sukzessive aus der Patientenbehandlung verschwinden, ausgenommen der (meist auch ohne placebokontrollierte randomisierte Studien) glücklicherweise gut etablierten chirurgischen Verfahren.

### Karriere-Bias

Ähnliche Auswirkungen hat der Karriere-Bias, auf den J. Kriz hinsichtlich des Begriffs der Wissenschaftlichkeit in der Psychotherapie verwies: „„Hunderte von Studien“, die oft als besonderes Qualitätsmerkmal propagiert werden, sprechen nicht unbedingt für die Brauchbarkeit von Therapiemethode X im Gesundheitssystem, sondern für die Brauchbarkeit dieser Methode im Rahmen universitärer Karrieremuster ... Es ist nicht so, dass die meisten Schulen auf den Nachweis der Psychotherapie-wirksamkeit verzichtet haben ... Vielmehr ist es so, dass die meisten Wissenschaftler auf die Untersuchung der Psychotherapie-wirksamkeit ... bei vielen Schulen verzichtet haben, deren theoretischen Konzepte nicht so gut in die

»publish or perish«-Strukturen der Forscherkarrieren passten.“ [4] – Das heißt, dass prioritär diejenigen Therapien erforscht werden, die derzeit *en vogue* sind und damit Vorteile der persönlichen Karriere des jeweiligen Forschers bieten. Dies führt zu dem Trugbild, dass diese Therapien, da mehr erforscht, auch therapeutisch sinnvoller und besser seien.

### Bias der großen Zahl

Unabhängig von der statistischen Power sollen RCTs eine große Patientenzahl (möglichst im dreistelligen oder höheren Bereich) umfassen, um die Zuverlässigkeit der Ergebnisse zu verbessern. Jedoch erfüllen meist nur wenige Patienten der zu untersuchenden Erkrankung die Einschlusskriterien von RCTs. Das bedeutet, dass RCTs hauptsächlich nur zu Massenerkrankungen durchgeführt werden können, und dass viele Studien an der Patientenrekrutierung scheitern. So hätten in einer Machbarkeitsstudie der Universitäts-Frauenklinik Heidelberg zur Mitteltherapie von den 1922 Patientinnen, die innerhalb von 28 Monaten mit Verdacht auf Mamma-Ca. operiert wurden, nur 29 (1,5%) der initial operierten (2% aller Frauen mit gesichertem Mamma-Ca.) in eine entsprechende Studie randomisiert werden können [5].

Mit begrenzten Patientenzahlen kann nur eine begrenzte Anzahl von Therapien evaluiert werden. Große RCTs verdrängen RCTs zu anderen Therapien, mit teils fatalen Konsequenzen: „I know that several of the larger firms use overpowered trials as a way of keeping competitors out of that particular subject.“ [6] Dieses Dilemma verhindert Studien oder führt zur Durchführung von RCTs in ärmeren Ländern und somit zu einer gesundheitswissenschaftlichen Neo-Kolonialisierung.

### Praxisferne randomisierter Studien

„Although the amount of medical evidence increases daily, many core questions remain unanswered.“ [1] –

RCTs geben auf viele Fragen der täglichen Patientenversorgung keine oder nur unzureichende Antworten. Neben dem oben schon genannten liegen die Gründe auch in folgendem: RCTs haben meist nur kurze Beobachtungszeiten (z. B. beträgt die Beobachtungszeit der meisten Antidepressivastudien 6 Wochen), während die betreffenden Therapien, gerade bei chronischen Erkrankungen, oft jahrelang eingesetzt werden und positive und negative Langzeitwirkungen (z. B. Steroide bei Kindern) zentral für den Therapieentscheid sind. RCTs schließen in der Regel hochselektierte Patientengruppen ein. So erfüllen beispielsweise nur 15% der ambulant behandelten Patienten mit Major Depression nach DSM-IV die Einschlusskriterien von 31 Antidepressiva-Studien aus fünf führenden Psychiatrie-Zeitschriften [7]. Von diesen 15% der Patienten spricht wiederum nur ein Teil auf die antidepressive Behandlung an, so dass der Nachweis der Wirksamkeit der antidepressiven Behandlung in RCTs nur für einen Bruchteil der Patienten gilt. In RCTs werden meist nur Patienten ohne bedeutsame Begleiterkrankung aufgenommen, in der Praxis müssen jedoch Patienten mit oft mehreren Begleiterkrankungen behandelt werden, wobei relevante Wechselwirkungen der jeweiligen Behandlungen auf die Zweit- und Dritterkrankungen zu erwarten sind. RCTs werden primär im akademischen Setting durchgeführt, das jedoch nur weniger als 1% der zu behandelnden Patienten abdeckt. Zwar werden gelegentlich RCTs auch in Praxen der Allgemeinversorgung durchgeführt, jedoch differieren hier wiederum Diagnostik, Therapie und Studienoutcome bzw. konkretes Behandlungsziel; ferner werden RCTs meist Diagnose-spezifisch durchgeführt, während die Behandlung beim Primärarzt im allgemeinen Symptom-spezifisch erfolgt, wo eine je ausgiebige Diagnostik weder realistisch noch sinnvoll ist [1]. Ein Resümee lautet deshalb: „30 years later, Cochrane’s approach clearly enhanced the number of published papers, but it has rather limited effect on clinical practice and policies – even when applying the most sophisticated electronic techniques.“ [1].

## Divergente Ergebnisse

Die Diskussion um Sinn oder Unsinn des Mammographiescreenings zeigt beispielhaft, dass verschiedene professionelle Evidenz-basierte Auswertungen identischer klinischer Studien zu verschiedenen Schlussfolgerungen und sogar zu entgegengesetzten Therapieempfehlungen kommen können [8]. Auch andere Beispiele zeigen, dass verschiedene professionelle Meta-Auswertungen derselben Studien völlig unterschiedliche Gesamtaussagen ergeben können: wirksam vs. unwirksam (z. B. [9]).

Aus der Divergenz ärztlicher Beurteilungen erfolgt ursprünglich die Forderung nach kontrollierten klinischen Studien. RCT-Ergebnisse sind aber ebenfalls oft divergent [10], weswegen systematische Reviews (Metaanalysen) von RCTs durchgeführt werden. Nun erweisen sich aber auch die Ergebnisse der systematischen Reviews von RCTs als divergent. Die Gründe sind, dass sowohl RCTs als auch systematische Reviews aufwendig, störanfällig und schwierig durchzuführen und zu beurteilen sind. Ob größere Formalisierung und Dimensionierung der Studien das Problem überwinden können, ist fraglich.

## Asymmetrische Beweiskraft

Während ein positives RCT-Ergebnis im Prinzip (bei optimaler Durchführung) eine Beweiskraft hat, gilt dies *nicht* für ein negatives RCT-Ergebnis. Dies ist *kein* valider Nachweis der Unwirksamkeit. („The major weakness of the randomized trial is the difficulty for protection against *false negativity*.“ [11]) – Diese grundsätzliche RCT-Asymmetrie ist in der heutigen Methodendiskussion weitgehend unbeachtet („Orthodoxy always invokes the danger of Type One errors to ensure the occurrence of Type Two errors.“ [12]), auch wenn Praktiker immer wieder darauf hinweisen (z. B. [9]). Eine Vielzahl an Fehlermöglichkeiten können falsch-negative Ergebnisse in RCTs provozieren, wie sich aus de-

taillierterer Analyse konkreter RCTs leicht zeigt: darunter fallen kompensatorische Zusatzbehandlung in der durch die Studienmedikation unterbehandelten Kontrollgruppe, suboptimale Wahl von Patienten, Ärzten, Therapiemodalitäten, Beobachtungszeiträumen, Erhebungsinstrumenten, etc.; ebenfalls kann das Ergebnis fälschlich nivelliert werden durch Gefälligkeitsauskünfte, mangelnde Differenzierungskraft der Erhebungsbögen, Gruppenangleichung, kognitive Interaktionen, Konditionierungseffekte, simplifizierende Studiendesigns, Intention-to-treat Analyse, u.a.m. (Übersicht Tabelle 2, weiteres z. B. in [13–15]). Während die experimentelle Forschung diesen Punkt durchaus kennt („As with most experimental situations, negative results are rarely conclusive, being attributable not only to basic flaws in the hypothesis but also to technical limitations or errors.“ [16]), werden in der klinischen Methodologie, wo der konservative Ansatz vorherrscht, d. h. die Sorge um Vermeidung falsch positiver Ergebnisse überwiegt, diese Negativ-Fehlermöglichkeiten kaum reflektiert. Während die formale Struktur einer perfekt

**Tabelle 2.** Faktoren, die in randomisierten Studien die Tendenz zu falsch negativen Ergebnissen fördern können.

- Therapiefehler, falsche Dosierung
- Zusätzliche und kompensatorische Behandlung
- Spezifisch wirksame „Placebo“-behandlung
- Drop-outs und Non-Compliers
- Kontamination und Intention-to-treat Analyse
- Informed Consent
- Gefälligkeitsauskunft, experimentelle Unterordnung
- Mangelnde Differenzierungskraft der Erhebungsmethode
- Tendenz zu mittelwertigen Angaben
- Gruppenangleichung
- Fallstricke bei der Patientenrekrutierung
- Konditionierungseffekte
- Kognitive Interaktionen
- Störung des Arzt-Patienten-Verhältnisses
- Fehlattribution
- Simplifiziertes Studiendesign (Mega-Studien)

durchgeführten RCT ein Schutz gegen falsch positive Ergebnisse ist, gibt es in der RCT keinen Formalismus zur Identifizierung von Falsch-negativ-Bias. Deshalb ist prinzipiell keine wirksame Therapie davor gefeit, infolge von formal perfekten, aber inhaltlich unachtsamen Studien als unwirksam geprüft zu werden und so aus der Patientenbehandlung zu verschwinden [13, 14].

## Ethische Probleme der randomisierten Therapiezuweisung

Eine RCT darf nur durchgeführt werden, wenn „equipoise“ besteht, d.h. wenn keine ausreichenden Gründe für die Überlegenheit der Prüftherapie sprechen. Diese ethische Voraussetzung des „equipoise“ scheint aber bei RCTs im allgemeinen nicht gegeben: So zeigte eine Übersichtsarbeit zu 945 publizierten und nicht-publizierten Studien, dass die Prüftherapie in 47,3% statistisch signifikant überlegen war, die Kontrolltherapie jedoch nur in 3,7% [17]. Ähnliche Ergebnisse zeigen auch andere Übersichten [13]. Das heißt: Es werden Patienten in Kontrollgruppen von RCTs mehrheitlich benachteiligt. Die Konsequenz ist, dass ein individual-ethisch orientierter Arzt streng genommen seine Patienten nicht in die Kontrollgruppe von RCTs einbringen dürfte.

## Therapiepräferenz

Eine entschiedene Präferenz bei Patienten und Ärzten pro oder kontra einer der Therapien verhindert oder erschwert die Teilnahme an RCTs. Klassische Beispiele sind Hausgeburt versus Klinikgeburt oder operative Entfernung einer lokal begrenzten Krebserkrankung (ja oder nein). In einer prospektiven Studie zu akuten Ohr- und Atemwegsinfektionen (anthroposophische vs. schulmedizinische Behandlung) wurden die Patienten befragt, ob sie zum Zwecke einer wissenschaftlichen Untersuchung einer randomisierten Therapiezuweisung zustimmen würden: 35% der schulmedizinischen, aber nur

3,2% der anthroposophischen Behandlungsgruppe waren zur Randomisation bereit [18]. Unter solchen Bedingungen sind RCTs schwerlich realisierbar. In manchen Ländern wird das Präferenzproblem dadurch umgangen, indem Patienten die Therapie außerhalb der Studien vorenthalten oder ihre Verfügbarkeit verheimlicht wird und sie so bei Therapiewunsch, insbesondere bei lebensbedrohlichen Erkrankungen, zur Studienteilnahme genötigt werden; dies ist jedoch ethisch problematisch [19].

## Mediokritätsbias

Manche medizinische Verfahren, vor allem in komplexen medizinischen Situationen, sind in hohem Maße abhängig vom individuellen Können und der Erfahrung des behandelnden Arztes oder des behandelnden Teams. Für solche Situationen können RCTs prinzipiell ungeeignet sein. Das Problem wurde kürzlich anlässlich einer im Jahr 2000 publizierten RCT zum Geburtsmodus bei Beckenendlage (BEL) – geplante vaginale Entbindung (VE) versus geplanter Kaiserschnitt – ausführlich diskutiert [20]. Jene Studie (121 Zentren, 26 Länder, 2088 Frauen) zeigte eine signifikant höhere neonatale Mortalität und Morbidität unter geplanter VE. Jedoch waren außerhalb dieser Studie, in erfahrenen Zentren, die Ergebnisse der vaginalen Entbindung weit besser. Auf folgende grundsätzlichen Probleme wurde hingewiesen [20]:

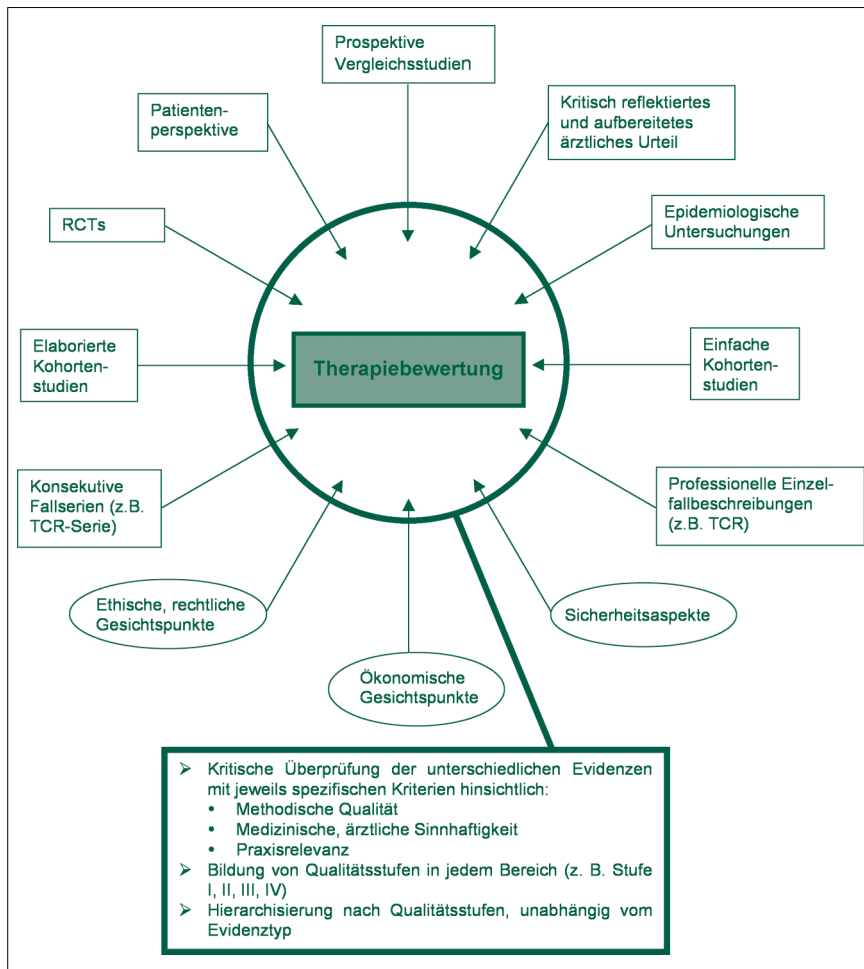
Die sichere VE bei BEL benötigt große Erfahrung der Geburtshelfer, sicheres klinisches Urteil und großes Können („skill“), und zwar hinsichtlich Geburtstechniken, differenzierter Diagnostik, Ultraschall, Patienteneinschätzung, Kindsüberwachung, Geburtsleitung. Wie auch in operativen Verfahren, so ist auch hier „skill“ der stärkster Prädiktor für das Ergebnis. Eine multizentrische RCT hingegen kann das hohe Erfahrungs- und Kompetenzniveau spezialisierter Zentren nicht zur generellen Richtlinie aller beteiligten Kliniken erheben; sie benötigt vielmehr einfache Leitlinien einer durchschnittlichen Patientenversorgung, Vereinfachung und

Nivellierung. Und so wurden auch Zentren, z. B. in USA, mit wenig Erfahrung in vaginaler Entbindung und entsprechend niedrigen Raten (13%), aus studentischen Gründen dazu verleitet, diese Raten auf wenigstens 40% zu erhöhen. Dies wurde erleichtert, da im Kontext der RCT die Haftungsrisiken vernachlässigbar waren. Dies hatte dramatische Folgen: Die RCT prüfte eine qualitativ mittelmäßige und damit unsichere vaginale Entbindung, die vaginale Entbindung erbrachte schlechtere Ergebnisse, mit weitreichenden Folgen für die Geburtshilfe weltweit, außerdem mit Folgen für erfahrene und befähigte Ärzte, die nun verunsichert sind und ihrem ärztlichen Urteil nicht mehr vertrauen.

Für solchen komplexen Fragen – in heterogenen Populationen, für komplexe, individuell zu variierende Therapieprozesse, für Beurteilungs- und Skill-abhängige Verfahren – haben RCTs aufgrund ihrer Notwendigkeit zu Vereinfachung und Homogenisierung eine intrinsische Limitierung. Die Bedingungen der sicheren und optimalen VE, die eigentlich erforscht werden müssten, wurden aus studentischen Gründen nivelliert und damit ausgeschaltet. Zur Erforschung und Verbesserung „skill“-abhängiger Verfahren bei komplexen Situationen würden sorgfältige historische Vergleiche und die Analyse der Techniken erfahrener Ärzte eine bessere Urteilsgrundlage bieten. „The philosophical limits of evidence based medicine include failing to appreciate and cultivate the complex nature of sound clinical judgement ...“ [20].

## Ausblick

Unter Idealbedingungen ist die RCT ein perfekt zuverlässiges Evaluationsmodell. Unter Realbedingungen jedoch ist die RCT zwar manchmal die Methode der Wahl, oft aber ist sie ungeeignet, unethisch, nicht realisierbar, nicht relevant oder führt zu unzuverlässigen Ergebnissen. Um den realen Gegebenheiten der medizinischen Versorgung gerecht zu werden, um Verzerrungen durch Kommerz-Bias, Karriere-Bias, Mediokritätsbias, Bias der großen Zahl



**Abb. 1.** Informationssynthese aus verschiedenen Arten von Evidenz, die auch den Einbezug des wissenschaftlich aufgearbeiteten ärztlichen Urteils ermöglicht.

zu minimieren, um ethische und strukturelle Probleme gering zu halten, um den Fortschritt der Medizin nicht unnötig zu erschweren [6, 22] und um die ärztliche Fremdbestimmung durch Fachfremde, die Instrumentalisierung der RCT durch Bürokraten zu vermindern [22], ergeben sich für die Zukunft folgende Notwendigkeiten:

- 1) Analyse der Kriterien des „sound clinical judgement“ und Professionalisierung des ärztlichen Urteils;
- 2) Aufhebung pauschaler Evidenzhierarchien;
- 3) Entwicklung eines Systems der Informationssynthese aus verschiedenen Arten der Evidenz (s. Abb. 1): neben RCTs auch andere prospektiv vergleichende Studien, einfache und elaborierte Kohortenstudien und konsekutive Fallserien, profes-

sionelle Einzelfallbeschreibungen (z. B. nach CBM, siehe Beitrag von Kiene, S. 301–306), kritisch reflektiertes und aufgearbeitetes ärztliches Urteil, hochwertige retrospektive Untersuchungen, epidemiologische Untersuchungen, Patientenperspektive. Alle Arten dieser Evidenzerhebungen müssen einer kritischen, jeweils spezifisch angepassten Kriterien-gestützten Qualitätsbeurteilung unterzogen werden, die medizinische und praxisrelevante Gesichtspunkte mitumfassen. Für jede einzelne Evidenzklasse könnten dann Qualitätsstufen definiert werden (z.B. Qualitätsstufe I, II, III, IV). Eine hierarchische Wertigkeit könnte, individuell für jede Fragestellung, nach methodischer Qualität sowie medizinischer Sinnhaftigkeit und praktischer Modellvalidität

der Evidenzen festgelegt werden, wobei dann Evidenzen der jeweils höchsten Qualitätsstufe gleichwertig in die Synthese eingehen.

## Literatur

- [1] De Maeseneer JM, van Driel ML, Green LA, van Weel C. The need for research in primary care. *Lancet* 2003; 362: 1314–9.
- [2] Holden C. NIH trial to test chelation therapy. *Science* 2002; 297:1109.
- [3] Horrobin DF. Innovation in the pharmaceutical industry. *J R Soc Med* 2000; 93:341-5.
- [4] Kriz J. Gutachten über den Begriff der Wissenschaftlichkeit in der Psychotherapie. *Zeitschrift des Schweizerischen Berufsverbandes für angewandte Psychologie, SBAP* 2003.
- [5] Gerhard I, Abel U, Loewe-Mesch A, Huppmann S, Kuehn JJ. Problematik randomisierter Studien in der Komplementärmedizin dargestellt am Beispiel der Misteltherapie bei Patientinnen mit Mammakarzinom. *Forsch Komplementärmed* 2004; 11:150–7.
- [6] Horrobin DF. Are large clinical trials in rapidly lethal diseases usually unethical? *Lancet* 2003; 361:695–7.
- [7] Zimmerman M, Mattia JI, Posternak MA. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *Am J Psychiatry* 2002; 159(3):469–73.
- [8] The mammography debate: A crisis for Evidence Based Medicine? In: 4. Symposium Evidenzbasierte Medizin. 14.–15. März 2003, Freiburg i.Br.: 2003.
- [9] Hopayian K. The need for caution in interpreting high quality systematic reviews. *Br Med J* 2001; 323:681–4.
- [10] Horwitz RI. Complexity and contradiction in clinical trial research. *Am J Med* 1987; 82:498–510.
- [11] Freireich EJ. The randomized clinical trial is not the best and certainly not the only way to conduct clinical research. *Advances: The Journal of Mind-Body Health* 1997; 13(2):41–4.
- [12] Eysenck HJ. Prediction of cancer and coronary heart disease mortality by means of a personality inventory: results of a 15-year follow-up study. *Psychol Rep* 1993; 72:499–516.
- [13] Kiene H. *Komplementäre Methodenlehre der klinischen Forschung. Cognition-based Medicine*. Berlin, Heidelberg, New York: Springer-Verlag, 2001.

- [14] Kiene H. Kritik der klinischen Doppelblindstudie. München: MMV Medizin Verlag, 1993.
- [15] Kienle GS, Kiene H. Die Mistel in der Onkologie - Fakten und konzeptionelle Grundlagen. Stuttgart, New York: Schattauer Verlag, 2003.
- [16] Herberman RB. Animal tumor models and their relevance to human tumor immunology. *J Biol Resp Modif* 1983; 2(3):217–26.
- [17] Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. *Controlled Clin Trials* 1987; 8(4):343–53.
- [18] Hamre HJ, Fischer M, Heger M, Riley D, Haidvogel M, Baars E et al. Anthroposophic vs. Conventional Therapy of Acute Respiratory & Ear Infections: A Prospective Outcomes Study. *Wien Klin Wochenschr* 2005; 117: 256–68.
- [19] Menikoff J. The hidden alternative: getting investigational treatments off-study. *Lancet* 2003; 361:63–7.
- [20] Kotaska A. Inappropriate use of randomised trials to evaluate complex phenomena: case study of vaginal breech delivery. *Br Med J* 2004; 329:1036–9.
- [21] Hannah ME, Hannah WJ, Hewson SA, Hodnett ED, Saigal S, Willan AR. Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial. *Lancet* 2000; 356:1375–83.
- [22] Kienle GS, Karutz M, Matthes H, Matthiessen PF, Petersen P, Kiene H. Konkurs der ärztlichen Urteilskraft? *Deutsches Ärzteblatt* 2003; 100(Heft 33):A 2142–6.

**Korrespondenzadresse:**

Dr. med. Gunver S. Kienle, Institut für angewandte Erkenntnistheorie und medizinische Methodologie e.V., Schauinslandstraße 6, D-79189 Bad Krozingen  
 Tel.: ++49-7633-806682  
 Fax: ++49-7633-806696  
 e-mail: [gunver.kienle@ifaemm.de](mailto:gunver.kienle@ifaemm.de)  
<http://www.ifaemm.de>

