

Vier Effekte, Phänomene und Paradoxe in der Medizin

Ihre Relevanz und ihre historischen Wurzeln

Peter Kleist

PFC Pharma Focus AG, Volketswil



Quintessenz

Als Leser von medizinischer Fachliteratur sollten Sie sich von den folgenden vier Phänomenen nicht aufs Glatteis führen lassen:

- Der Hawthorne-Effekt: Ändern Studienteilnehmer aufgrund der Studienteilnahme ihr natürliches Verhalten, kann das – insbesondere in der Kontrollgruppe – zu einer Überschätzung von Behandlungseffekten führen.
- Das Regression-to-the-mean-Phänomen: Werden Patienten auf der Basis sehr hoher bzw. tiefer Ausgangswerte in eine Studie eingeschlossen, sieht die statistische Variabilität von Folgemessungen wie eine Verbesserung durch die Therapie aus.
- Das Simpson-Paradox: Wenn (un)bekannte Einflussfaktoren (Confounder) auf das Studienergebnis einwirken, kann das Gesamtergebnis einer (Fallkontroll-)Studie durch Subgruppenanalysen auf den Kopf gestellt werden.
- Das Will-Rogers-Phänomen: Verbessern sich die diagnostischen Möglichkeiten oder wird die Prävalenz einer Erkrankung künstlich angehoben, kann sich die Prognose eines Patienten verbessern, ohne dass sich an seinen Messwerten irgendetwas geändert hat.

Die Durchführung randomisierter und kontrollierter Studien mit einer exakten Festlegung des Vorgehens durch das Protokoll ist das erfolgreichste Mittel zur Ausschaltung dieser vier Phänomene.

Summary

Four effects, phenomena and paradoxes in medicine. Their relevance and historical roots

The reader of medical literature must beware of the following four phenomena:

- *Hawthorne effect: involvement in a study may change a study participant's normal behaviour and result in overestimation of treatment effects, particularly in the control group.*
- *Regression to the mean: when patients are enrolled in a study with extremely high or low values, the statistical variability of repeat measurements may look like a true treatment effect.*
- *Simpson's paradox: if an (un)known confounder influences the result of a (case control) study, subgroup analyses may reverse the effects shown in the overall study results (comparison of cases and controls).*
- *Will Rogers phenomenon: a patient's mean prognosis may improve merely due to more sensitive diagnostic tools or an artificial increase in disease prevalence, without any change in the patient's measurement values.*

The conduct of randomised and controlled studies on the basis of a specified methodological protocol is the most successful means of eliminating or neutralising these four phenomena.

Einführung

Daten lügen nicht – oder etwa doch? Zumindest verschleiern sie ab und zu die Wahrheit. Dass unterschiedliche Erscheinungsformen eines Bias die Ergebnisse klinischer Studien verzerren, ist Ihnen wahrscheinlich bewusst. Dieser Beitrag möchte Sie mit der klinischen Bedeutung und dem historischen Hintergrund von vier ausgewählten und vielleicht weniger bekannten «Merkwürdigkeiten» vertraut machen, die uns bei der Interpretation von klinischen Daten in die Irre führen können: dem Hawthorne-Effekt, dem Regression-to-the-mean-Phänomen, dem Simpson-Paradox und dem Will-Rogers-Phänomen.

Der Hawthorne-Effekt

Der Hawthorne-Effekt besagt folgendes: Die Teilnehmer einer Studie ändern ihr natürliches Verhalten, weil sie wissen, dass sie an einer Studie teilnehmen und unter Beobachtung stehen. Der Hawthorne-Effekt wurde beispielsweise gezielt an zwei Patientengruppen untersucht, die sich einem chirurgischen Eingriff am Knie unterzogen: Diejenigen Patienten, die über die Teilnahme an einer wissenschaftlichen Untersuchung zum Befinden nach dem Eingriff informiert wurden, fühlten sie sich nach der Operation generell besser und berichteten über deutlich weniger postoperative Schmerzen als jene Patienten, die sich dessen nicht bewusst waren [1].

Eine vergleichbare Studie wurde an Patienten mit milden bis moderaten Tumorschmerzen durchgeführt. Diese erhielten in einem Crossover-Design Naproxen bzw. Placebo. Wussten die Patienten, dass sie Teilnehmer an einer Studie waren, entfaltete sogar die Placebogabe subjektiv stärkere analgetische Effekte als die Verabreichung von Naproxen bei jenen Patienten, welche die Schmerzmittelgabe im Rahmen der Standardbehandlung zu erhalten glaubten, das heisst ohne Kenntnis einer Studienteilnahme [2]. Auch die in weiteren Untersuchungen ermittelte individuelle Reaktion auf Hypnotika, Koffein oder

nikotinhaltige Kaugummis bzw. auf korrespondierende Placebos hing massgeblich davon ab, ob die Patienten über die Studienteilnahme informiert waren oder nicht [3].

Innerhalb einer klinischen Studie betrifft der Effekt prinzipiell alle Patienten, das heisst sowohl die in der Interventions- als auch die in der Kontrollgruppe. Wenn die in der Kontrollgruppe einer Studie erzielten Ergebnisse durch einen Hawthorne-Effekt überdurchschnittlich gut sind, kann das den Nachweis eines Therapieeffektes in der Interventionsgruppe erschweren oder sogar unmöglich machen, weil der Spielraum für das Aufzeigen der Überlegenheit einer Massnahme sehr klein wird – zum Beispiel für Zahnpflegemittel, da alle Studienpatienten eine im Vergleich zu den Alltagsbedingungen verbesserte Mundhygiene betreiben [4]. In bezug auf die generelle Patientenführung ist dieser Effekt aber nicht nur unerwünscht, denn alle Patienten profitieren ja – allein aufgrund der Teilnahme an einer Studie.

Der Hawthorne-Effekt betrifft jedoch nicht nur Versuchspersonen, sondern auch die Versuchsleiter: So führte die Mitwirkung an Beobachtungsstudien zu einer signifikant restriktiveren Verschreibung von Antibiotika bei Kindern mit Atemwegsinfekten [5] bzw. zu signifikant weniger Spitalüberweisungen von Kindern mit Asthma [6], jeweils aufgrund eines stärker evidenzorientierten Managements der Erkrankung durch die Studienärzte. In einer randomisierten multizentrischen Studie an niereninsuffizienten Patienten zeigte ein spezifisches Anämiebehandlungsprotokoll keine Vorteile gegenüber der lokalen Standardtherapie – und zwar aufgrund einer im Verlauf der Studie kontinuierlich verbesserten Einstellung der Patienten in der Vergleichsgruppe [7]. Und gegenüber Patienten ausserhalb einer Studie wiesen Kontrollpatienten in Untersuchungen zur Postinfarktthrombolysie sogar bessere Überlebensraten auf, die durch den Einschluss von Patienten mit einer besseren Prognose allein nicht zu erklären sind [8].

Fazit

Der Hawthorne-Effekt kann zu einer Überschätzung der Wirksamkeit in den Kontrollgruppen von klinischen Studien führen und den Nachweis der Wirksamkeit einer effektiven Therapie erschweren oder sogar verhindern. Nichtspezifizierte Standardbehandlungen in Vergleichsgruppen kontrollierter Studien oder im Rahmen von Beobachtungsstudien sind oftmals nicht repräsentativ für eine Standardbehandlung unter Alltagsbedingungen.

Historische Randbemerkungen

Der Hawthorne-Effekt wurde durch Untersuchungen in den Hawthorne-Werken der Western Electric Company in den USA bekannt [9]. Die zwischen 1924 und 1927 durchgeführten Stu-

dien sollten zeigen, ob einfache Veränderungen der Arbeitsbedingungen Auswirkungen auf die Produktivität der Arbeiter haben. Die Produktivität stieg, um ein Beispiel zu nennen, sowohl unter verbesserten als auch unter verschlechterten Beleuchtungsverhältnissen an – und selbst dann, wenn «alte» Glühbirnen durch neue mit identischer Leuchtstärke ausgewechselt wurden.

Entscheidend für die Ergebnisse waren offenbar nicht die äusseren Gegebenheiten, sondern die Teilnahme an der Studie an sich, die in diesem Zusammenhang erfahrene Aufmerksamkeit und der Eindruck der Versuchsteilnehmer, dass irgendetwas verändert wurde.

Das Regression-to-the-mean-Phänomen

Das Regression-to-the-mean-Phänomen spielt dann eine Rolle, wenn Messungen grundsätzlich keine konstanten Werte ergeben, sondern eine Variabilität aufweisen, etwa aufgrund intraindividuelle Schwankungen, situativer Anpassungen oder wegen Messungenauigkeiten.

Ein Beispiel hierfür sind Blutdruckwerte, die um einen unbekanntem Mittelwert streuen. Die Variabilität der Messungen – sowohl bei einem einzelnen als auch konsekutiv bei einer Gruppe von Patienten – unterliegt einer Normalverteilung. Nun erfolgt die Auswahl von Studienpatienten häufig auf der Basis extremer, das heisst deutlich erhöhter oder sehr tiefer Ausgangswerte. Diese unterliegen keiner Normalverteilung mehr, sondern repräsentieren das extreme Ende der Variabilität in der untersuchten Patientenpopulation. Wäre der «wahre» Wert bekannt gewesen, hätte man wohl einen Teil dieser Patienten nicht in die Studie aufgenommen. Die im Studienverlauf vorgenommene Wiederholungsmessung bei Patienten, die aufgrund einmal gemessener hoher Blutdruckwerte in eine Studie eingeschlossen wurden, ergibt mit grosser Wahrscheinlichkeit Werte, die näher am (unbekanntem) individuellen Mittelwert bzw. am Gruppenmittelwert liegen, also Werte, die im Vergleich zum Ausgangswert tiefer sind [10].

Ein hypothetisches Beispiel soll dies veranschaulichen: Im Rahmen eines Screenings für eine klinische Studie wird bei 250 Hypertonikern der diastolische Blutdruck gemessen. Der Gruppenmittelwert der normalverteilten Messwerte liegt bei 90 mm Hg. Eine Wiederholungsmessung würde wiederum einen Gruppenmittelwert im Bereich von 90 mm Hg ergeben, denn die bei einigen Patienten – im Vergleich zur ersten Messung – individuell höheren und bei anderen Patienten individuell tieferen Werte gleichen sich im Gesamtkollektiv weitgehend aus.

Da das Protokoll der geplanten Studie jedoch nur Patienten erfassen möchte, deren diastolischer Wert über 95 mm Hg liegt, können von den

ursprünglichen 250 Patienten letztlich nur 80 in die Studie aufgenommen werden. Der Mittelwert der jetzt nicht mehr normalverteilten Werte dieses ausgewählten Patientenkollektivs liegt bei 100 mm Hg. Sehr wahrscheinlich wurden bei einem beträchtlichen Anteil dieser 80 Patienten «hohe Ausreisserwerte» erhoben.

Leitet man nun bei den 80 Studienpatienten eine antihypertensive Therapie ein und misst nach vier Wochen erneut den Blutdruck, liegt der mittlere diastolische Blutdruck jetzt vielleicht bei 95 mm Hg, also 5 mm Hg tiefer als der Ausgangswert. Aber: Hätte man bei diesen Patienten die Wiederholungsmessung vor der Therapieeinleitung durchgeführt, wäre ebenfalls eine Senkung des Blutdrucks nachweisbar gewesen, und ein Teil der Patienten hätte das Einschlusskriterium von >95 mm Hg wohl gar nicht mehr erfüllt.

Die zu beobachtende Senkung des Blutdrucks ist nämlich lediglich Ausdruck eines statistischen Phänomens, und zwar des Regression-zum-Mittelwert- bzw. des Regression-to-the-mean-Phänomens, welches die natürliche Variabilität von Wiederholungsmessungen wie eine tatsächliche Verbesserung aussehen lässt [11]. Je weiter ein (Ausreisser-)Wert vom Mittelwert der Verteilung abweicht, desto höher ist die Wahrscheinlichkeit, dass eine Folgemessung näher bei diesem Mittelwert liegt. Damit lässt sich auch erklären, warum die Wirkung von Antihypertensiva scheinbar parallel zum Schweregrad der Hypertonie zunimmt. Je höher die Ausgangswerte, desto stärker die Blutdrucksenkung – ein Effekt, der jedoch in gleicher Weise unter Placebo nachweisbar ist: So betrug in der «Australian Mild Hypertension Study» die mittlere diastolische Blutdrucksenkung unter Placebo in der Patientengruppe mit den höchsten Ausgangswerten über 9 mm Hg [12].

Die Bedeutung des Regression-to-the-mean-Phänomens ist nicht zu unterschätzen und kann in nahezu allen Bereichen der klinischen Medizin auftreten. So besteht bei der (kleinen) Gruppe von Patientinnen, bei denen – trotz effektiver Therapie – im ersten Behandlungsjahr die Knochendichte abnimmt, im Folgejahr eine hohe Wahrscheinlichkeit für eine Zunahme der Knochendichte, ohne die Therapie zu ändern und allein aufgrund des besagten Phänomens [13]. Die markante Verbesserung ausgeprägter allergischer Symptome in einer offenen Beobachtungsstudie lässt sich nicht allein auf die Behandlung mit einem Antihistaminikum zurückführen, sondern ist sicherlich auch Ausdruck des Regression-to-the-mean-Phänomens [14]. Und die Wirksamkeit einer Impfung gegen Meningitis C (Reduktion von bis zu 90% der Meningitisfälle in Grossbritannien) wird mit Sicherheit überschätzt, da die Impfung zu einer Zeit erfolgte, in der die Inzidenz besonders hoch war [13, 15].

Fazit

Unkontrollierte Beobachtungsstudien an ausgewählten Patienten mit hohen/tiefen Ausgangswerten überschätzen die tatsächlichen Therapieeffekte – wichtig ist daher die gleichzeitige Untersuchung einer Kontrollgruppe. Regression-to-the-mean-Effekte können dann durch die Subtraktion der Wirkungen von Test- und Vergleichsbehandlung neutralisiert werden. Die unter einer Behandlung mit Placebo zu beobachtenden Verbesserungen sind – zumindest teilweise – auch auf Regression-to-the-mean zurückzuführen [16]. Der Einschluss von Patienten in klinische Studien sollte zudem auf der Basis multipler Ausgangsmessungen erfolgen, wenn mit einem Regression-to-the-mean-Phänomen zu rechnen ist.

Historische Randbemerkungen

Der Ausdruck «regression to the mean» geht auf den englischen Genetiker und Statistiker Sir Francis Galton (1822–1911) zurück, Halbcousin von Charles Darwin und «Erfinder» der Fingerprint-Identifikation. Bei Messungen der Körperlänge von über 900 erwachsenen Kindern und deren Eltern stellte er nämlich fest, dass sehr grosse Eltern in der Regel Kinder mit einer im Vergleich zu ihnen geringeren Körperlänge haben (die aber immer noch grösser als der Durchschnitt ist), während sehr kleine Eltern im allgemeinen Kinder haben, die grösser als sie selbst sind (aber immer noch kleiner als der Durchschnitt) [17].

Das Simpson-Paradox

Dies ist ein Phänomen in Verbindung mit unterschiedlichen statistischen Auswertungsansätzen. Je nachdem, ob die Gesamtgruppe oder Untergruppen von Studien betrachtet werden, erhält man einander widersprechende Ergebnisse. Das Paradox tritt hauptsächlich bei nicht-randomisierten Beobachtungsstudien auf (insbesondere in Fallkontrollstudien), lässt sich aber auch im Rahmen von nichtvaliden Berechnungen der «number needed to treat» (NNT) in Metaanalysen beobachten.

Rauchen ist vorteilhaft für unsere Gesundheit. Zu dieser Annahme könnte man gelangen, wenn man seinen Überlegungen die Beziehung zwischen dem Rauchen und der Langzeitüberlebensrate bei über 1300 britischen Frauen zugrundelegen würde. Zwischen 1972 und 1974 wurden die Frauen befragt, ob sie rauchen oder nicht. Zwanzig Jahre später schaute man, ob die Betroffenen noch lebten – mit folgendem Resultat: Gut 31% der Nichtraucherinnen, aber nur 24% der Raucherinnen waren in der Zwischenzeit verstorben.

Bei dieser Betrachtungsweise übersieht man jedoch den Einfluss des Lebensalters. Erst in den

1960er und 1970er Jahren begannen Frauen zu rauchen, und es waren vor allem Jüngere. Die anteilmässig viel kleinere Gruppe von Raucherinnen war daher auch noch ungleich auf verschiedene Altersgruppen verteilt. Werden die Daten nun in Abhängigkeit vom Alter analysiert, dann findet man – von einer Ausnahme abgesehen – in allen Altersgruppen eine höhere Sterblichkeit unter den Raucherinnen [18].

Ein weiteres Beispiel aus einer Fallkontrollstudie soll dies verdeutlichen [19]: Im Rahmen eines historischen Vergleichs wurden die Erfolge zweier Verfahren zur Nierensteinentfernung miteinander verglichen: jene der offenen Chirurgie (Daten von 1972–1980) und jene der perkutanen Nephrolithotomie (Daten von 1980–1985). Die offene Chirurgie wies eine Erfolgsrate von 78% auf (273 von 350 Patienten), die – scheinbar effektivere – Nephrolithotomie eine solche von 83% (289 von 350 Patienten). Wertet man die Daten jedoch in Abhängigkeit von der Steingrösse (< bzw. ≥ 2 cm) aus, wird das Resultat auf den Kopf gestellt: Bei kleinen Steinen war die offene Chirurgie gegenüber der Nephrolithotomie erfolgreicher (93 vs. 83% bzw. 81/87 vs. 234/270 Patienten), bei grossen allerdings auch (73 vs. 69% bzw. 192/263 vs. 55/80 Patienten). Rechnen Sie ruhig nach – es stimmt. Diese Umkehrung der Ergebnisse nennt man Simpson-Paradox [20].

Was zeigen uns diese Beispiele? Bei den Auswertungen der Gesamtgruppen blieb ein wichtiger Einflussfaktor (Confounder) auf die Resultate (Lebensalter der Frauen bzw. die Auswahl eines Behandlers durch den behandelnden Arzt in Abhängigkeit von der Steingrösse) unberücksichtigt. Hinzu kommt, dass sich die Grösse der analysierten Untergruppen und die dort beobachteten Prävalenzen (Anteil an Raucherinnen) bzw. Inzidenzen (Erfolgsraten der Nierensteinentfernung) deutlich voneinander unterscheiden. Aus diesem Grund ist es auch nicht möglich, im Rahmen einer Metaanalyse die NNT auf der Basis der Summe aller Ereignisse zu berechnen, wenn die Gruppengrössen und die Ergebnisse der Einzelstudien stark voneinander abweichen. Dies soll ein weiteres Beispiel verdeutlichen: Während die einer Metaanalyse zugrundeliegenden Einzelstudien überwiegend einen Vorteil der institutionalisierten Raucherentwöhnung gegenüber Programmen ohne personelle Unterstützung zeigten, betrug die Gesamtrate der Patienten, die mit dem Rauchen aufhörten, in den Interventionsgruppen 14,3% (546/3820 Patienten), in den Kontrollgruppen hingegen (höhere) 15,5% (356/2301 Patienten) [21].

Fazit

Unbekannte Einflussfaktoren können die Ergebnisse von Beobachtungsstudien, insbesondere von Fallkontrollstudien, verfälschen. Hohe Fallzahlen und statistische Signifikanzen bieten zudem keinen Schutz vor falschen Schlussfolge-

rungen [22]. Liegt ein Simpson-Paradox vor, kennen wir letztlich die Wahrheit nicht: Denn auch die Untergruppenauswertung kann durch eine weitere, unbekanntere Variable beeinflusst werden, die – falls berücksichtigt – erneut das Ergebnis ins Gegenteil verkehren würde. Daher sollten, sofern möglich, randomisiert-kontrollierte Studien durchgeführt werden. Bei publizierten Metaanalysen ist auf die Methode zur Berechnung der NNT zu achten.

Historische Randbemerkungen

Der Name des Simpson-Paradoxes geht auf den amerikanischen Statistiker Edward Hugh Simpson zurück, der diese Erscheinung 1951 beschrieb [23]; dies allerdings zu unrecht, denn schon lange zuvor hatten der britische Mathematiker Karl Pearson (1899) und der schottische Statistiker George Udny Yule (1903) bereits auf dieses statistische Problem aufmerksam gemacht [24].

Der Bekanntheitsgrad des Paradoxes stieg in den 1970er Jahren stark an, als die Universität in Berkeley beschuldigt wurde, Frauen bei der Zulassung zum Studium zu benachteiligen. Tatsächlich wurden 44% der männlichen Bewerber, aber nur 34% der Frauen angenommen. Eine detaillierte Auswertung auf Fakultätsebene zeigte jedoch, dass in fast allen Bereichen anteilmässig mehr Frauen als Männer zugelassen wurden. Wo finden wir den Confounder? Frauen bewarben sich vorzugsweise für solche Fächer, bei denen die Ablehnungsquote insgesamt hoch war, während sich Männer hauptsächlich für die Fächer interessierten, bei denen viele Studenten zugelassen wurden [25].

Das Will-Rogers-Phänomen

Kann sich die Prognose eines Patienten verbessern, ohne dass sich an seinem Gesundheitszustand oder seinen Messwerten irgendetwas geändert hat? Ja – und zwar aufgrund des Will-Rogers-Phänomens bzw. des «stage migration effects». Statistiker sprechen manchmal auch von «krimineller (Daten-)Vereinigung».

Als medizinisches Phänomen wurde es erstmals in der Onkologie beschrieben: Eine amerikanische Studie verglich zwei Kohorten von Patienten mit einem Lungenkarzinom miteinander. Die jüngere Kohorte aus den 1970er und 1980er Jahren wies eine bessere Sechs-Monats-Überlebensrate auf als die ältere aus den 1950er und 1960er Jahren. Beiden Kohorten lag die gleiche Stadieneinteilung des Karzinoms (TNM-Klassifikation) zugrunde. Die scheinbare Prognoseverbesserung war jedoch nicht auf eine Therapieoptimierung zurückzuführen, sondern auf neue bildgebende Verfahren, die sich zwischenzeitig etabliert hatten, und die eine sensitivere Erfassung von Metastasen ermöglichten. Dieser Um-

stand sorgte dafür, dass ein Teil der Patienten in der jüngeren Kohorte jeweils einem höheren Erkrankungsstadium zugeordnet wurden, als dies in der älteren Kohorte der Fall war. Als Folge davon verbesserte sich die Prognose sowohl in den unteren Krankheitsstadien (weil die Patienten mit einer schlechteren Prognose in das nächsthöhere migrierten) als auch die im fortgeschrittenen Stadium (weil die hochgestuften Patienten eine durchschnittlich bessere Prognose als diejenigen Patienten aufwiesen, die diesem Stadium leichter zugeordnet werden konnten) [26]. Vor kurzem publizierte Populationsstudien haben das Phänomen der scheinbaren Prognoseverbesserung – bedingt durch eine optimierte Diagnostik und eine Stadienmigration – ebenfalls für das Prostatakarzinom [27] und das kolorektale Karzinom [28] gezeigt.

Das Will-Rogers-Phänomen spielt auch dann eine Rolle, wenn sich durch die grosszügigere Definition einer Erkrankung deren Prävalenz ändert. Die Herabsetzung von als pathologisch angesehenen Schwellenwerten der Plasmaglukose, des Blutdrucks und des Cholesterins sorgt für eine Ausweitung der Population von Diabetikern, Hypertonikern und Patienten mit Lipidstoffwechselstörungen. Allein durch die Hinzunahme von Patienten mit einer günstigeren Prognose verbessert sich die durchschnittliche Prognose aller Patienten in der Gesamtgruppe [29].

Fazit

Trauen Sie keinen Untersuchungen mit historischen Kontrollgruppen. Fortschritte in der Diagnostik oder eine artifizielle Erhöhung der Erkrankungsprävalenz können eine Prognoseverbesserung vortäuschen. Diese darf dann nicht einer in der Zwischenzeit erreichten Verbesserung der Therapiestandards oder neu auf den Markt gekommenen Arzneimitteln zugeschrieben werden [30].

Empfohlene Literatur

- Bland JM, Altman DG. Regression towards the mean. *BMJ*. 1994;308:1499.
- Julious SA, Mullee MA. Confounding and Simpson's paradox. *BMJ*. 1994;309:1480–1.
- Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med*. 1985;312:1604–8.

Historische Randbemerkungen

Der Namensgeber für das soeben beschriebene Phänomen war der amerikanische Humorist und Entertainer Will Rogers. Zur Zeit der Wirtschaftskrise in den 1930er Jahren wanderten viele Einwohner Oklahomas (die sog. «Okies») nach Kalifornien aus. Rogers, selber 1879 in Oklahoma geboren, hielt wenig von den Auswanderern, aber noch weniger von den Kaliforniern, und ihm wird der folgende Ausspruch zugeschrieben [31]: «When the Okies left Oklahoma and moved to California, they raised the average intellectual level in both states.»

Schlussbemerkungen

Die Interpretation klinischer Daten ist manchmal nicht einfach und das Offensichtliche spiegelt nicht immer die Wahrheit wider. Als Leser von medizinischer Fachliteratur sollten Sie mit potentiellen Fussangeln vertraut sein und mit einem kritischen Auge über die Studienergebnisse und die daraus gezogenen Schlussfolgerungen wachen. Insbesondere unkontrollierte Beobachtungsstudien, Fallkontrolluntersuchungen, historische Vergleiche und Vergleichsgruppen ohne standardisierte Vorgehensweisen verfügen über Schwachstellen – ein wesentlicher Grund dafür, dass wir medizinische Evidenz, so weit wie ethisch vertretbar und praktikabel, auf randomisierte kontrollierte und prospektive Studien mit einer eindeutigen Festlegung der Vorgehensweisen abstützen müssen.

Da nicht alles Publierte erforderlichen Qualitätsansprüchen genügt, steigt mitunter die «NNR», das heisst die «number (of studies) needed to read», um eine klinische Fragestellung valide beantworten zu können.

Korrespondenz:
Dr. med. Peter Kleist
PFC Pharma Focus AG
Chriesbaumstrasse 2
CH-8604 Volketswil
peter.kleist@pfc.ch

- Fey MF, Bühler A. Klinische Fachliteratur kritisch lesen. Teil I: Fussangeln in Fallberichten und Fallserien. *Schweiz Med Forum*. 2001;1(7):161–5.

Das vollständige Literaturverzeichnis [1–31] finden Sie in der Onlineausgabe dieses Artikels unter www.medicalforum.ch/pdf/pdf_d/2006/2006-46/2006-46-194.pdf.

Vier Effekte, Phänomene und Paradoxe in der Medizin

Ihre Relevanz und ihre historischen Wurzeln

Peter Kleist

PFC Pharma Focus AG, Volketswil

Literatur

- 1 De Amici D, Klersy C, Ramajloji F, Brustia L, Politi P. Impact of the Hawthorne effect in a longitudinal clinical study: the case of anesthesia. *Control Clin Trials*. 2000;21:103–14.
- 2 Bergmann JF, Chassany O, Gandiol J, Deblois P, Kanis JA, Segrestaa JM, et al. A randomized clinical trial of the effect of informed consent on the analgesic activity of placebo and naproxen in cancer pain. *Clin Trials Meta-Anal*. 1994; 29: 41–7.
- 3 Kaptchuk TJ. The double-blind, randomized, placebo-controlled trial. Gold standard or golden calf? *J Clin Epidemiol*. 2001; 54: 541–9.
- 4 Winston JL, Bartizek RD, McClanahan SF, Mau MS, Beiswanger BB. A clinical methods study of the effects of triclosan dentifrices on gingivitis over six months. *J Clin Dent*. 2002;13:240–8.
- 5 Mangione-Smith R, Elliott MN, McDonald L, McGlynn EA. An observational study of antibiotic prescribing behavior and the Hawthorne effect. *Health Serv Res*. 2002;37:1603–23.
- 6 Zinman R, Bethune P, Camfield C, Fitzpatrick E, Gordon K. An observational asthma study alters emergency department use: the Hawthorne effect. *Pediatr Emerg Care*. 1996;12:78–80.
- 7 Brimble KS, Rabbat CG, McKenna P, Lambert K, Carlisle EJ. Protocolized anemia management with erythropoietin in hemodialysis patients: a randomised controlled trial. *J Am Soc Nephrol*. 2003;14:2654–61.
- 8 Jha P, Deboer D, Sykora K, Naylor CD. Characteristics and mortality outcomes of thrombolysis trial participants and nonparticipants: a population-based comparison. *J Am Coll Cardiol*. 1996;27:1335–42.
- 9 Roethlisberger FH, Dickson WJ. Management and the worker. Cambridge, MA: Harvard University Press: 1939.
- 10 Bland JM, Altman DG. Regression towards the mean. *BMJ*. 1994;308:1499.
- 11 Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol*. 2005;34:215–20.
- 12 The Australian therapeutic trial in mild hypertension. Report by the management committee. *Lancet*. 1980;i:1261–7.
- 13 Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. *BMJ*. 2003;326:1083–4.
- 14 Bachert C, Vovolis V, Margari P, Murrieta-Aguttes M, Santoni JP. Mizolastine in the treatment of seasonal allergic rhinoconjunctivitis: a European clinical experience with 5408 patients managed in daily practice. *Allergy*. 2001;56:653–9.
- 15 Wise J. News roundup: meningitis C rates show steep fall. *BMJ*. 2001;322:70.
- 16 Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med*. 2001;344:1594–602.
- 17 Galton F. Regression towards mediocrity in hereditary stature. *J R Anthropol Inst*. 1886;15:246–63.
- 18 Appleton DR, French JM, Vanderpump MPJ. Ignoring a covariate: an example of Simpson's paradox. *Am Stat*. 1996;50:340.
- 19 Charig CR, Webb DR, Payne SR, Wickham OE. Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy. *BMJ*. 1986;292:879–82.
- 20 Julious SA, Mullee MA. Confounding and Simpson's paradox. *BMJ*. 1994;309:1480–1.
- 21 Cates CJ. Simpson's paradox and calculation of number needed to treat from meta-analysis. *BMC Med Res Methodol*. 2002;2:1. Epub 2002 Jan 25.
- 22 Heydtmann M. The nature of truth: Simpson's paradox and the limits of statistical data. *QJM*. 2002;95:247–9.
- 23 Simpson EH. The interpretation of interaction in contingency tables. *J R Statist Soc B*. 1951;2:238–41.
- 24 Tu Y-K, West R, Ellison GTH, Gilthorpe MS. Why evidence for the fetal origins of adult disease might be a statistical artefact: The "reversal paradox" for the relation between birth weight and blood pressure in later life. *Am J Epidemiol*. 2005;161:27–32.
- 25 Bickel PJ, Hjammel EA, O'Connell JW. Sex bias in graduate admissions: data from Berkeley. *Science*. 1975; 187: 398–404.
- 26 Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med*. 1985;312:1604–8.
- 27 Albertsen PC, Hanley JA, Barrows GH, Penson DF, Kowalczyk PD, Sanders MM, Fine J. Prostate cancer and the Will Rogers phenomenon. *J Nat Cancer Inst*. 2005;97:1236–7.
- 28 Shahrier M, Ahnen DJ. Colorectal cancer and survival in Europe: the Will Rogers phenomenon revisited. *Gut*. 2000;47:463–4.
- 29 Schwartz LM, Woloshin S. Changing disease definitions: Implications for disease prevalence. Analysis of the third national health and nutrition examination survey, 1988–1994. *Eff Clin Pract*. 1999;2:76–85.
- 30 Fey MF, Bühler A. Klinische Fachliteratur kritisch lesen. *Schweiz Med Forum* 2001;1(7):161–71.
- 31 Vineis P, McMichael AJ. Bias and confounding in molecular epidemiological studies: special considerations. *Carcinogenesis*. 1998;19:2063–7.

Korrespondenz:

Dr. med. Peter Kleist
PFC Pharma Focus AG
Chriesbaumstrasse 2
CH-8604 Volketswil
peter.kleist@pfc.ch