

Über den Schatten springen

Die investigative Radiologie muss sich der Bedrohung ihrer Arbeit durch systematische Fehler stellen



Prof. Werner Albert Golder

Korrespondenzadresse

Prof. Werner Albert Golder
Radiologie, Association d'Imagerie Medicale, 65 rue Raymond
Poincaré, 10000 Troyes, France
Tel.: ++ 33/3/25 73 27 58
Fax: ++ 33/3/25 82 60 43
werner.golder@orange.fr

Bibliografie

DOI <https://doi.org/10.1055/s-0043-123828>
Online-Publikation: 28.12.2017
Fortschr Röntgenstr 2018; 190: 235–236
© Georg Thieme Verlag KG, Stuttgart · New York
ISSN 1438-9029

Viele neue diagnostische und therapeutische Verfahren führen nach einer von Offenheit geprägten und Optimismus getragenen Test- und Einführungsphase in der breiten Anwendung zu mehr oder weniger enttäuschenden Ergebnissen. Die zunehmende Verfeinerung und Qualitätskontrolle klinischer Studien, die für die Prüfungen durch das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) und den Gemeinsamen Bundesausschuss (G-BA) und die Überführung von Innovationen in erstattungsfähige Leistungen angesichts begrenzter finanzieller Ressourcen essenziell sind, haben diesen Trend bisher nicht verlangsamt, geschweige denn gestoppt. Dass selbst nach strengen Kriterien konzipierte und realisierte Studien zu Ergebnissen führen können, die Nutzen vortäuschen, wo tatsächlich kein positiver Effekt entsteht oder sogar Nachteile eintreten, hängt in vielen Fällen damit zusammen, dass das Problem des systematischen Fehlers, der auch in der deutschsprachigen Literatur häufig als Bias bezeichnet wird, zu wenig beachtet wird.

Systematische Fehler können die innerhalb einer klinischen Studie optimierten Bedingungen des Methodenvergleichs nachhaltig stören, indem sie Ergebnisse produzieren, die in einer bestimmten Richtung von den wahren Werten abweichen. Dadurch werden Unterschiede gezielt vergrößert, verkleinert oder sogar invertiert. Die Kenntnis der Quellen des Bias und die Auswahl der geeigneten Maßnahmen zur Verringerung ihres Einflusses sind für klinische Studien von grundsätzlicher Bedeutung. Untersuchungen an Patienten und Probanden sind ohnehin mit so vielen Unwägbarkeiten verbunden, dass man sie nicht mehr als unvermeidbar durch systematische Fehler zusätzlich belasten sollte. Die in der Radiologie dominierenden Fall-Kontroll-, Kohorten- und Interventionsstudien sind für viele systematische Fehler anfällig und zugleich nur mit Mühe vor ihnen zu schützen bzw. gegen

sie zu verteidigen; die Kenntnis der wichtigsten Formen des Bias ist zwingend geboten.

Trotz sorgfältiger Planung und Durchführung lassen sich systematische Fehler in klinischen Studien grundsätzlich nicht vermeiden [1]. Das Bemühen, sie zu verringern, ist aber ein zentrales Qualitätsmerkmal und mitbestimmend für die Verlässlichkeit der Ergebnisse. Wenn ein Bias methodisch nicht ausgeschaltet werden kann, müssen seine potenziellen Auswirkungen auf das Ergebnis abgeschätzt werden – und zwar so exakt wie möglich. Dazu bedarf es der gezielten Suche, der Identifizierung und der Bestimmung der Richtung, in die der Fehler führt. Doch während andere Einflussgrößen wie Zufall und Störfaktoren quantifiziert und durch gezielte rechnerische Korrekturen nachträglich abgeschwächt werden können, fehlt diese Option für die systematischen Fehler. Systematische Fehler können auch nicht durch die Vergrößerung des Stichprobenumfangs eliminiert werden. Dieses Konzept eignet sich nur für die Reduktion von Zufallsfehlern. Die systematischen Fehler müssen in der Planungsphase der Studie bekämpft werden. Randomisierung, und Verblindung sind die wichtigsten Instrumente der Abwehr. Die Randomisierung fördert die biasfreie Verteilung der bekannten, aber auch unbekannter Einflussgrößen zwischen den verschiedenen Gruppen. Die Verblindung reduziert die systematischen Fehler, die aus der Kenntnis der Vorgeschichte der Teilnehmer resultieren.

- Es gilt als unstrittig, dass die in vielen Publikationen dokumentierten hohen Werte für die Parameter der statistischen Genauigkeit dadurch mitbedingt sind, dass die Auswertung von Spezialisten/Subspezialisten und nicht von Ärzten, die überwiegend in der Allgemeinversorgung ihres Faches tätig sind, durchgeführt werden. Dieses Bias kann man dadurch abschwächen, dass die Messdaten parallel von Fachvertretern,

die mit dem Forschungsprojekt nicht unmittelbar verbunden sind, ausgewertet werden.

- Die homogene Zusammensetzung von Studienkollektiven hat nicht nur Vorteile. Untersuchungen, in deren Rahmen Patienten mit voll ausgeprägter Krankheit und Gesunde miteinander verglichen werden, sind für das Spektrum der Patienten in der Praxis weniger repräsentativ als solche, in die auch Patienten mit geringeren Beschwerden sowie Kontrollpersonen mit anderen, z. T. verwechselbaren Krankheiten eingehen.
- Es ist ein Kennzeichen qualitativ hochwertiger Studien, wenn an die Auswahl und Gewinnung des Referenzstandards bei allen Teilnehmern hohe Ansprüche gestellt werden. Im Idealfall führen die Untersucher das aktuelle Prüfverfahren und den Referenzstandardtest ohne Kenntnis des Resultats der jeweils anderen Methode durch. Manchmal muss aber aus praktischen Gründen ein weniger gut geeigneter Parameter zum Referenzstandard gemacht werden und manchmal wird selbst er nicht bei allen Teilnehmern komplett erhoben und in die Auswertung eingebracht. Auf eine derartige Fehlerquelle muss man bei allen Vergleichsstudien diagnostischer Tests achten.
- Viele Untersucher sind davon überzeugt, dass das Design einer Studie nur dann ideal ist, wenn die Erkrankung nach strengen Kriterien verifiziert wird. Tatsächlich sind aber Studien, in denen nicht definitiv bestätigte Fälle ausgeschlossen werden, mit einem systematischen Prüffehler belastet. Wenn sich nur Patienten mit eindeutig positiven Testresultaten der Prüfung des Referenzstandards unterziehen, kann die Sensitivität falsch zu hoch eingeschätzt werden.
- Bei einer Gruppe von Patienten mit einheitlicher Diagnose formuliert man identische Befunde großzügiger als in einer heterogenen Population. Um die Bestimmung der Genauigkeit eines Tests nicht zu verzerren, sollten die Untersucher daher grundsätzlich Stichproben verwenden, in denen die Prävalenz des Symptoms bzw. der Erkrankung der in der klinisch relevanten Bevölkerung entspricht. In der Studienpraxis ist es aber oft schwer, diese Forderung zu erfüllen. Die große Zahl gleichartiger Bilddokumente und Fragestellungen bei der Auswertung radiologischer Studien ist jedenfalls nicht typisch für die Prävalenz derartiger Materials in der Versorgungspraxis selbst spezialisierter Einrichtungen.
- Der Einfluss von Lernkurven auf die praktische Anwendung neuer Techniken führt sowohl auf der Seite der Studienärzte als auch auf jener der Testpersonen über den gesamten Beobachtungszeitraum hinweg zu einer Verfälschung der Ergebnisse. Dabei sind vor allem zwei Mechanismen bedeutsam. Zum einen kann beim Vergleich von zwei Verfahren in ein und derselben Hand die neue, aktuell beforschte Methode gegenüber der Kontrolltechnik einen Vorteil genießen, wenn man Anfängerfehler und kleine Mängel toleriert. Zum anderen kann die Leistungsfähigkeit unterschätzt werden, wenn man die Möglichkeiten von Hard- und Software aus Unkenntnis oder Zögern nicht vollständig ausschöpft.
- Je sensitiver ein Test ist, desto länger überlebt scheinbar die Population, an der er durchgeführt wird. Die Folgen der Vorverlegung des Diagnosezeitpunkts verzerren vor allem die Ergebnisse von Studien zur Effektivität und Effizienz von Früh-

erkenntnisuntersuchungen maligner Tumoren. Nur wenn der Test und die daraus gezogenen Konsequenzen das Überleben tatsächlich verlängern, unterscheiden sich die Überlebenszeiten der getesteten und der nicht getesteten Personen um mehr als die sogenannte Vorlaufzeit. Sonst führt die vorverlegte Diagnose nur dazu, dass die getesteten Personen um diese Zeitspanne länger wissen, dass sie krank sind, aber nicht dazu, dass sie länger als die Vorlaufzeit am Leben bleiben.

- In der Regel werden neue diagnostische/therapeutische Verfahren von Vertretern jener Disziplinen erprobt, die sie auch später hauptverantwortlich einsetzen. Dieses Privileg birgt jedoch die grundsätzliche Gefahr, dass das Urteil aus verständlichem Eigeninteresse zu positiv ausfällt. Umgekehrt können negative oder warnende Stimmen überwiegen, wenn Vertreter einer oder mehrerer konkurrierender Disziplinen die Prüfung durchführen. Durch derartige professionelle Vorurteile bedingte Verzerrungen lassen sich reduzieren, wenn die Studien interdisziplinär durchgeführt werden oder zumindest die Auswertung im fachübergreifenden Team stattfindet.

Viele systematische Fehler, von denen klinische Studien bedroht werden, lassen sich auch bei großer Planungssorgfalt nicht ausschalten, sondern nur verringern. Umso wichtiger ist es, die Verzerrungen im Detail zu analysieren und die dabei gewonnenen Ergebnisse vollständig zu dokumentieren, auch und gerade, um den durch Studien erzielbaren Erkenntnisgewinn zur Basis erstattungsfähiger medizinischer Leistungen zu machen. Das Biasproblem muss im Rahmen jeder wissenschaftlichen Studie gesondert und umfassend untersucht und dargestellt werden. Meistens addieren sich mehrere Biases zu einem veritablen Zerrbild. Die Bewertung der Rolle des Bias als Alternativerklärung für eine beobachtete Verknüpfung ist bei der Interpretation jedes Studienergebnisses unverzichtbar, auch wenn der wissenschaftliche Ertrag dabei sinkt. Untersuchungen, die von systematischen Verzerrungen weitgehend frei sind, führen nahezu ausnahmslos zu statistisch weniger signifikanten Resultaten als biasbelastete Studien. Die Untersucher sollten alle potenziellen Verzerrungen ihrer Arbeit und die Maßnahmen, die sie zu deren Ausschaltung getroffen haben, beschreiben, um bei den Adressaten die Voraussetzungen dafür zu schaffen, sich selbst ein umfassendes Bild von der möglichen Beeinflussung der Befunde und Schlussfolgerungen durch systematische Fehler zu machen [2]. Diese Analyse sollte einen selbstständigen Abschnitt im Diskussionsteil jeder wissenschaftlichen Veröffentlichung darstellen und von den Herausgebern bei der Einreichung als obligatorischer Bestandteil des Manuskripts gefordert werden. Andeutungen oder summarische Hinweise genügen nicht. Arbeiten, die keine befriedigende Erklärung zum Biasproblem enthalten, sollten nicht zur Veröffentlichung angenommen werden.

Literatur

- [1] Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; 32: 51–63
- [2] Bossuyt PM, Reitsma JB, Bruns DE et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138: W1–W12