Prof. Dr. med. Dr. h.c. Peter Paul Nawroth

„Grilling the data“: Plädoyer für einen intensiveren Blick auf die Ergebnisse klinischer Studien

Zusammenfassung

Einleitung: Ein Arzt möchte seinem Patienten helfen. Die auf hippokratische Traditionen zurückgeführte Grundeinstellung des Arzt-Seins, das ‚primum non nocere‘, wird durch eine zweite Aufforderung ergänzt: ‚secundum cavere‘ = vorsichtig sein. Dies wird oft nur als Aufforderung verstanden, sich genau das Problem des Patienten anzuschauen. Eine andere Interpretation lautet: eine Verpflichtung zum tiefgründigen Verstehen dessen, was man tut. Das gelingt nur durch eigenständiges Lesen und Interpretieren der Originalliteratur. Modern formuliert: ‚kritisch erkunden, denkend verantworten‘.

Methoden: Dafür reicht es nicht, sich auf Kongressen fortzubilden, Leitlinien zu kennen oder der Werbung des Pharmavertreters zuzuhören. Um dem ‚kritisch Erkunden, denkend verantworten‘ = ‚cavere‘ als Vorgehensweise gerecht zu werden, muss man willens und in der Lage sein, die Originalarbeiten zu verstehen und die Interpretation der Wissenschaftler zu hinterfragen.

Ergebnisse: Dies sollte eine wichtige ärztliche Tätigkeit sein. Sie darf nicht alleine Lobbyisten und Fachgesellschaften überlassen werden, denn der eigenständig hinterfragende Blick des klinisch tätigen Arztes auf die Originaldaten führt häufig zu anderen Schlüssen als der der Wissenschaftler, Experten und Fachgesellschaften. Wenn viele Ärzte wenigsten einige Originalstudien lesen, wird dies eine Wirkung erzeugen, der die Experten und Interessenvertreter sich nicht entziehen können: Die aus Patientensicht klinisch relevanten Kriterien werden die Kriterien ersetzen, die aus berufspolitischer Sicht oft zur Überinterpretation der von ihrer eigenen Forschung enthusiasmierten Wissenschaftler führen.

Diskussion: Die Chance für eine Aufwertung des Arztberufes und Verbesserung der Patientenversorgung wird an einem Beispiel beschrieben.

Schlüsselwörter

Fortbildung, kritischer Rationalismus, Statistik, Arzt, Patientenversorgung

Crossref/doi

http://doi.org/10.24945/MVF.04.26.1866-0533.2852

Primum non (oder nil) nocere – nicht schaden, ist ein Grundsatz, der im Zentrum eines moralischen ärztlichen Handelns steht. Liest man das Zitat vollständig, eröffnen sich zusätzliche Perspektiven: „Primum non nocere, secundum cavere, tertium sanare.” – „Erstens nicht schaden, zweitens vorsichtig sein, drittens heilen.“ Es ist unerheblich, dass dieses Zitat nicht exakt so in den Schriften des Hippokrates von Kos steht und ob die lateinische Formulierung von dem römischen Arzt Scribonius Largus am Hof von Kaiser Tiberius oder dem Engländer Thomas Sydenham stammt. Zuerst geht es um eine normative Grundeinstellung: Auf keinen Fall schaden. Dann die Vorgehensweise: genau hinschauen. Sicherstellen, dass das ärztliche Tun einer Person wahrscheinlicher nutzt, als dass es schadet. Erst an dritter Stelle das Ziel: Heilen. Für das ‚secundum cavere‘ reicht die Vergewisserung dessen, was dem Patienten fehlt, nicht aus. Das „Bona diagnosis, bona curatio” stimmt, ist aber angesichts der Möglichkeiten evidenzbasierter-Medizin zu einseitig. Der Vorsicht ausübende Arzt kann diese Norm nur erfüllen, wenn er bereit ist, die Originaldaten der Wissenschaft mit dem Blick auf das Patientenwohl selber zu lesen und kritisch zu hinterfragen. Ein Arzt wird dem ‚secundum cavere‘ nicht gerecht, wenn er gutgläubig den Aussagen der Experten vertraut, die Fortbildungen halten und Leitlinien verfassen. Die gleiche Skepsis gilt es auch gegenüber Lobbyisten von Fachgesellschaften oder der Werbung von Firmen zu praktizieren. Der Arzt, der die Norm der Vorsicht auch auf dieser Ebene erfüllt, die dafür notwendigen Kenntnisse und den Blick auf das Patientenwohl in die Diskussion einbringt, erreicht zweierlei: Erstens eine Vertiefung und erhöhte Glaubwürdigkeit des Arzt-Patienten-Gesprächs und zweitens eine Änderung der Vorgehensweise der Experten und Leitlinienautoren, der Vortragenden bei Kongressen und Fortbildungen, derjenigen, die über Zulassung und Erstattung entscheiden. Modern formuliert lautet diese Forderung an den Arzt: ‚secundum cavere‘ = ‚kritisch erkunden, denkend verantworten.‘

>> Ist die Individualisierung der normativ begründeten ärztlichen Sorgfaltspflicht, die weit über die Verantwortungsübernahme durch Befolgen von Leitlinien und Fortbildungen hinausweist, für einen klinisch tätigen Arzt lebbar? Niemand kann neben der praktischen Tätigkeit alle für ihn relevanten Studien lesen und über deren Interpretation nachdenken. Dennoch wird der Effekt dramatisch sein, wenn jeder von uns nur einmal die Woche eine Originalarbeit liest und aus dem Blickwinkel des für seinen Patienten verantwortlichen Arztes auf die Daten inklusive Supplements schaut: Er wird sich nicht mehr mit dem zufriedengeben, was ihm Experten, nicht selten im Konzert mit Firmen und berufspolitischen Lobbyisten, vortragen. Die Folge: Die Experten, Firmen und Lobbyverbände werden sich den neuen Anforderungen anpassen, solange die zentrale Forderung lautet: Der Blick auf Studiendaten sollte unter dem Aspekt des Patientenwohls stattfinden. Wegfallen können für die Beurteilung einer patientenrelevanten klinischen Wirksamkeit uninteressanten p-Werte, die Nennung der relativen Risikoreduktion (ohne die Nennung der absoluten), Argumente, die Surrogatparameter nutzen, denn diese sind keine klinisch relevanten Endpunkte und auf Korrelationen beruhende Vermutungen, denn diese sollten nur als Hypothesen genannt werden. Wirksamkeit kann nur beurteilt werden, wenn z. B. in einer Studie die zeitliche Verzögerung des Endpunktes, Berechnung der Effektstärke, die Darstellung der Non-Responder etc. aufgeführt sind.

Im Folgenden wird ‚nur‘ an einem Beispiel (andere sind anderswo ausgeführt 1–2) aufgezeigt, wie überraschend anders und für Arzt und Patienten ebenso wie für die Wissenschaft bereichernd ein offenerer, patientenzentrierter Blick auf altgewohnte Daten sein kann.

Methoden und Analysemethoden: „Grilling the data“

Wang und Mitarbeiter veröffentlichten mit dem Titel „Grilling the data“ eine Arbeit (3), in der sie die Daten aus dem National

Health and Nutrition Examination Survey 2007 bis 2014 (NHANES Studie) einmal anders, als sonst in der Ernährungsepidemiologie üblich (4–5) analysierten. Mit dem für uns Ärzte ungewohnten Ansatz der „Spezifikationskurvenanalyse" konnten sie den Fehler umgehen, der bei der Auswertung von Beobachtungsstudien sonst den Lesern verborgen bleibt: die verschiedenen Vorannahmen, die Wissenschaftler machen, um ein bestimmtes Ergebnis zu erhalten. Diese verschiedenen Annahmen, die das Ergebnis einer Auswertung beeinflussen (3, 6–9), kann man mittels der Spezifikationskurvenanalyse offenlegen. Man erhält deswegen keine punktförmig genaue Antwort, sondern für jede der plausiblen Annahmen ein eigenes Ergebnis. Der Vorteil dieser Methode: Besser als durch das Konfidenzintervall erkennt der Leser statt eines Punktes der Pseudo-Gewissheit, einen Korridor, der die verschiedenen Möglichkeiten, die hinter den Daten stehen, abbildet. Die Sache ist in der Tat viel komplexer, als Wissenschaftler vereinfachend schlussfolgern: „Faktor Y erhöht das Risiko für eine Erkrankung um X Prozent und man kann viele Leben retten, wenn man dieses schädliche Verhalten reduziert".

Erkrankungen, die viele Risikoprädiktoren besitzen, z. B. kardiovaskuläre Erkrankungen, Demenz oder Krebs (eine Korrelation mit einer Erkrankung erlaubt nur von Prädiktoren zu sprechen, nicht von kausal verursachenden Faktoren), sind in Beobachtungsstudien nur schwer zu erforschen. Das liegt an den vielen Störfaktoren und deren gegenseitiger Beeinflussung. Wenn Menschen, die mehr rotes Fleisch essen, mehr wiegen und öfter rauchen, bedeutet das nicht, dass sie wegen des roten Fleisches rauchen und mehr wiegen. Es gibt mathematische Methoden, diese Interdependenzen einzuberechnen, aber dennoch bleibt das Ergebnis nur eine Korrelation, die abhängig vom gewählten Modell ist. Daher überrascht es nicht, dass man z. B. in Interventionsstudien zwar das Auftreten von Diabetes um fast 50% senken kann, aber nicht schwerwiegende kardiovaskuläre Endpunkte.

Zudem gibt es für jeden der eingepreisten Risikoprädiktoren zwei Probleme: Erstens, die Ergebnisse der Basisstudien haben ein recht großes Konfidenzintervall, und zweitens wird nicht berücksichtigt, dass z. B. ein 5-prozentiger Anstieg des Gewichtes das kardiovaskuläre Risiko mehr erhöht, als die Gewichtsabnahme in Interventionsstudien das Risiko senkt (siehe den Unterschied zwischen dem kardiovaskulären Risiko bei Zunahme des Körpergewichts und der viel geringeren absoluten Risikoreduktion unter Semaglutid). Das alleine sollte ein Grund sein, überoptimistischen Modellierungen und Simulationen nicht zu trauen! Ein weiteres Problem ist, dass Ernährungsgewohnheiten mit Bildung, sozialem Status, psychosozialen und anderen Faktoren verknüpft sind, deren gegenseitige Beeinflussung niemand mit Genauigkeit kennt und die aus methodischen Gründen nicht alle in einer Studie analysiert werden können. Das führt wieder zu dem Problem der nicht ausgeschalteten Störfaktoren, die vor allem dann ins Gewicht fallen, wenn der Effekt auf klinisch relevante Endpunkte (wie bei den meisten Ernährungs- und Lebensstilstudien) nur gering ist. Zuletzt muss bedacht werden, dass die zugrunde liegenden Daten aus Beobachtungsstudien (öfter als angenommen) gefälscht sein können, wie eine kürzlich erschienene, noch nicht begutachtete Studie aus Birmingham für Schlaganfälle berichtet.

Zudem werden auf großen Kohorten basierende Modellierungen der großen Heterogenität von Menschen nicht gerecht: Die Unterschiede bei der Verstoffwechselung von Glukose, Lipiden, Vitaminen etc. sind weitaus größer als die Unterschiede zwischen Fast Food und mediterraner Kost. Das erklärt, warum mathematisch korrekt gerechnet nicht unbedingt bedeutet, dass das Ergebnis der klinischen Überprüfung standhält.

Diskutiert man dies mit den Autoren großer Kohortenstudien, kommt das Argument: Man kann der Politik nur Ratschläge geben, wenn man bevölkerungsbasierte Studien mit einer möglichst genauen und umfassenden Simulation zur Verfügung stellt. Was ist mit kleinen, aber gut gemachten Interventionsstudien zur Überprüfung der klinischen Wirksamkeit von in großen Kohorten erzielten Annahmen? Sind Ärzte und Politiker wirklich auf eine Vereinfachung der Studieninterpretationen angewiesen?

Das als breiter Korridor dargestellte Ergebnis einer Spezifikationsanalyse (Abbildung 1) ergibt nach Offenlegung aller plausibler Vorannahmen einen abgewogenen Blick auf die Frage, ob rotes Fleisch schadet oder nicht:

Die Abbildung zeigt, dass rotes Fleisch in mehr Berechnungen nicht schadet oder sogar mehr nutzt, als dass es schadet. Doch Achtung: Die Mehrheit der Möglichkeiten bestimmt nicht die Wahrheit!

Nur weil mehr Berechnungen kein erhöhtes oder ein erniedrigtes Risiko ergeben, bedeutet das nicht, dass es keine Gefahr von rotem Fleisch geben könnte. Richtig wäre es daher, die Wissenschaftler

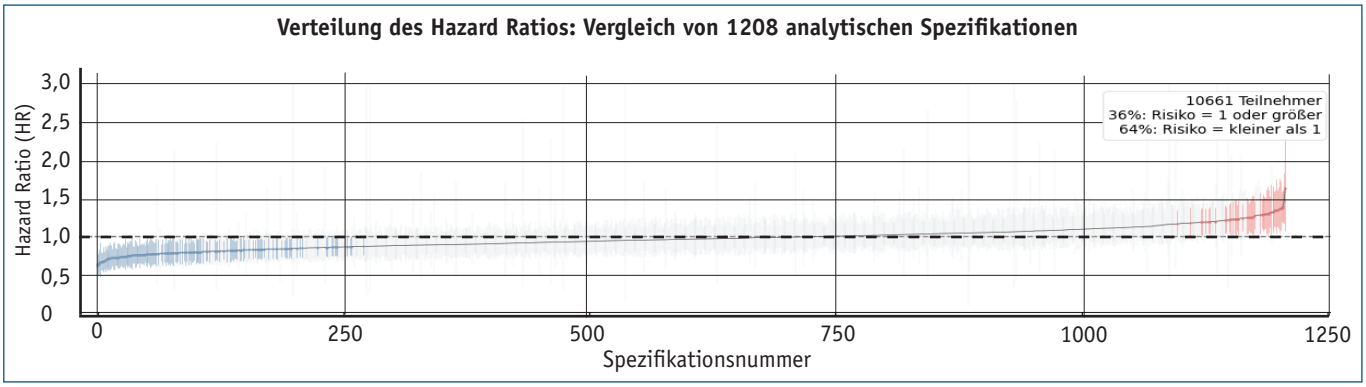


Abb. 1: Beispiel eines breiten Korridors einer Spezifikationsanalyse. Legende: Die roten Datenpunkte, die oberhalb der 1-er Linie liegen, weisen auf ein erhöhtes Risiko hin, zum Beispiel durch den Genuss roten Fleisches, die unterhalb der 1 auf ein verringertes Risiko. Abb. modifiziert nach (3) Wang et al.. (2024) Grilling the data: application of specification curve analysis to red meat and all-cause mortality. J Clin Epidemiol. Fig. 2.

würden sich für eine Förderung weiterer Studien einsetzen, um Subgruppen zu identifizieren, denen rotes Fleisch entweder nutzt oder schadet. Daraus ist erkennbar, dass Wissenschaftler den Fortschritt der Wissenschaft hemmen, dadurch dass sie unberechtigt feste Aussagen tätigen, statt offene Fragen anzusprechen. Offensichtliche Möglichkeiten für einen Erkenntnisfortschritt wären der Vergleich zwischen denen, die ein basal erhöhtes LDL-Cholesterin oder einen im oberen Normbereich liegenden Glukosewert haben, oder als Sahnehäubchen: Wie wäre es zu untersuchen, ob diejenigen, bei denen das LDL, die Glukose oder Marker der DNA-Schädigung (10) nach Verzehr von rotem im Vergleich zu weißem Fleisch besonders stark ansteigen, diejenigen sind, die mehr Herzinfarkte, Krebs und einen früheren Tod bei Konsum von rotem Fleisch erleiden? Solche Daten wären sinnvoller für ein Arzt-Patienten-Gespräch als der Wissenschaft übersimplifizierende Satz: ‚Rotes Fleisch könnte Ihnen schaden.'

Diese Feststellung gilt nicht nur für rotes Fleisch, sondern viele andere hochgelobte Präventionsmaßnahmen. Bedenkt man, dass selbst erfahrene Wissenschaftler sich oft nicht einig sind, welches die richtige und beste Methode der Datenanalyse ist (3–28), wird deutlich, dass wir (durch zu geringes kritisches Rückfragen?) uns an übersimplifizierte wissenschaftliche Aussagen gewöhnt haben. Wenn man dann noch bedenkt, dass in 62% der Ernährungsstudien keine Validierung der Qualität der Ernährungsanamnese nachzulesen ist und Abweichungen von Referenzmethoden über 400% betragen können (22), dabei die gefundenen Unterschiede oft nicht die 10%-Marke überschreiten, dann versteht man, dass es große Zweifel an der Wertigkeit vieler Ernährungsstudien geben muss (23–26). Könnte nicht die Darstellung des breiten Korridors der Möglichkeiten, anstatt der effekthascherischen Sensationsmeldung das Misstrauen gegen die auf epidemiologischen Beobachtungen beruhenden klinischen Empfehlungen beseitigen?

Diskussion

Schadet oder nutzt die Offenlegung der Ungewissheit der praktischen Medizin?

Nur auf den ersten Blick widerspricht die Datendarstellung als Korridor der Aufgabe der medizinischen Wissenschaft, dem praktisch tätigen Arzt Handlungsanweisungen zu geben. Das ‚secundum cavere' zerstört aus mehreren Gründen nicht den Anspruch des ‚tertium sanare':

1. Man soll von einer Wissenschaft nicht mehr erwarten, als ihre Methode und ihr Modell hergeben. Sonst gefährdet man den Patienten, da das auf dem Wissen um Wissenslücken basierende Abwägen einer trügerischen Sicherheit weicht. Von Beobachtungsstudien kann man nur Korrelationen erwarten, die auf einer Kausalität beruhen können oder auch nicht. Egal wie exquisit und komplex man rechnet.
2. Der breite Korridor der Möglichkeiten eröffnet dem Arzt die Chance, individuell zu entscheiden und mit seinem Patienten dort nach individuellen Lösungen zu suchen, wo – wie so oft – keine für den einen Patienten passenden Studiendaten vorliegen. Das macht den Wert des Arztes im Vergleich zu Doktor Google aus.
3. Das ‚cavere' bedeutet für den Arzt, sich Studien so genau anzusehen, dass er nicht gutgläubig oder gar leichtfertig Dateninterpretationen traut und diese unkritisch über seine Patienten ausgießt.
4. Durch sein eigenständiges Lesen stimuliert der praktisch tätige Arzt die verantwortlichen Wissenschaftler und Experten, nicht dort Gewissheit vorzugaukeln, wo durch die eingesetzte Methodik bedingt diese Gewissheit nicht vorliegt.

Schlussfolgerungen:

Das eigenständige kritische Lesen – und sei es nur vereinzelter Studien – ist angebracht: Wer hat nicht schon einmal den Verdacht in sich gespürt, dass es unter den etablierten Wissenschaftlern verschiedener Fachgebiete und damit auch unter den Gutachtern eine Übereinstimmung gibt, wie man Daten auswertet, obwohl diese einseitige Form der Auswertung nur einen kleinen Ausschnitt aus dem Korridor verschiedener gleichermaßen korrekter Ergebnisse darstellt (27–32)? Welche Rolle spielen dabei die Verlage, die darauf aus sind, Manuskripte zu publizieren, die viel zitiert werden und ein lebhaftes Presseecho und Werbeeinnahmen auslösen? Diese Schwächen des heute gelebten Peer-Review-Systems werden noch dadurch verstärkt, dass sich immer mehr subspezialisierte Journals herausbilden, die die Gutachter viel zu oft von den entsprechenden Fachgesellschaften empfohlen bekommen. Das Peer-Review-System verkommt nicht selten zu einer gegenseitigen Begutachtung von Wissenschaftlern mit gleicher Meinung und ähnlichen Interessen (31–37). Diese Schwächen kennt jeder, aber wie stellen wir uns auf diese Schwächen so ein, dass wir dem ‚cavere' ebenso gerecht werden, wie dem ‚non nocere', ohne das ‚sanare' zu verlieren?

Darauf gibt es folgende Antworten:

1. Durch Ertüchtigung, selber Originalstudien zu lesen und zu interpretieren, damit wir uns nicht von voreingenommenen Experten, die Teil der jeweiligen Szene sind, an der Nase herumführen lassen. Ein Ziel: Schnell erkennen, warum 85% der publizierten Studien in einer Lancet-Serie als „Müll" bezeichnet wurden (30–36).
2. Durch Änderung der Ausbildung (unter Lösung der bestehenden Copy-Right-Probleme, damit im Studentenunterricht Originaldaten gezeigt werden können): Zusätzlich zum Statistikkurs sollte während aller klinischer Semester das gemeinsame Lesen der Originaldaten Routine werden.
3. Durch das Bestehen auf patientenrelevanten Kriterien in Publikationen (z. B. Ausmaß der Endpunktverzögerung, Charakterisierung der Non-Responder, Berechnung der Effektstärke statt des p-Wertes, Spezifikationskurvenanalyse – wo angebracht, Analyse von Subgruppen …).
4. Durch das Pochen auf eine Offenlegung des Korridors der verschiedenen Möglichkeiten erkennen wir nicht nur die Unschärfe der gewonnenen Erkenntnis, sondern auch den künftigen Forschungsbedarf – ein Gewinn für Patienten, Wissenschaftler und die Glaubwürdigkeit der Wissenschaft.
5. Durch das Einüben mit mehr Ungewissheit zu leben (statt einer Leitlinien-Krankenkassen genehmer Kochbuchmedizin), wird die ärztliche Verantwortung aufgewertet, die Re-Akademisierung des Arztberufes eingeläutet, die Dialogfähigkeit erhalten (z. B. mit dem IQWIG).
6. Durch eine Übernahme des etliche Milliarden Euro schweren Marktes der wissenschaftlichen Publikationen durch europäische

Universitäten (siehe Stockholm Declaration) würden alle Zeitschriften die gleiche Formatierung verlangen, würden zigtausende Arbeitsstunden bei Mitarbeitern eingespart und unsinnige Publikationsanreize (37–38) abgeschafft.

7. Durch die Forderung, dass klinisch relevante Studien nicht hauptsächlich von Firmen gefördert werden, sondern dass die Möglichkeit staatlicher Studienförderung drastisch verstärkt wird, damit die offenen Fragen, die aus den von Firmen vorgelegten Studien hervorgehen, unabhängig und methodisch einwandfrei, mit patientenrelevanten Endpunkten untersucht werden. Diese Forschungsförderung wird weniger kosten als die Unsummen, die durch nicht adäquate Medikation ausgegeben werden. Konsequent angewandte und geförderte evidenzbasierte Medizin reduziert die Kosten weitaus mehr als jedes Sparpaket, das den fehlgeleiteten Inhalt nicht korrigiert.

Der Auftrag lautet also: Jeder von uns sollte einen Beitrag dazu leisten, dass die ungelösten Fragen offengelegt werden, damit die Wissenschaft die Möglichkeit erhält, durch neue Versuche das Morgen besser als das Heute zu gestalten. Dabei gilt es, eine Voraussetzung zu wahren: Die Induktion, also aufgrund einer Beobachtung auf die Verallgemeinerung zu schließen, ist ohne Überprüfung anhand anderer Fälle kein Werkzeug für den Erkennt-

Literatur

1. Nawroth P.: Nackte Medizin. Bloßstellung einer ideologisierten Wissenschaft. Achgut Edition, 2025
2. Nawroth P: Gebt der Medizin ihren Sinn zurück. Aufruf zu einer radikalen Umkehr im Gesundheitssystem. Springer 2018.
3. Wang Y, Pitre T, Wallach JD, de Souza RJ, Jassal T, Bier D, Patel CJ, Zeraatkar D. Grilling the data: application of specification curve analysis to red meat and all-cause mortality. J Clin Epidemiol. 2024 Apr;168:111278. doi: 10.1016/j.jclinepi.2024.111278. Epub 2024 Feb 12. PMID: 38354868; PMCID: PMC12289182.
4. Zeraatkar D, Guyatt GH, Alonso-Coello P, Bala MM, Rabassa M, Han MA, et al. Red and processed meat consumption and risk for all-cause mortality and cardiometabolic outcomes. Ann. Intern. Med. 2020;172:511e2.
5. An Pan 1, Qi Sun 1, Adam M Bernstein 1, Matthias B Schulze 1, JoAnn E Manson 1, Meir J Stampfer 1, Walter C Willett 1, Frank B Hu Red Meat Consumption and Mortality: Results from Two Prospective Cohort Studies. Arch Intern Med. 2012 Mar 12;172(7):555–563. doi: 10.1001/archinternmed.2011.2287
6. Tierney BT, Anderson E, Tan Y, Claypool K, Tangirala S, Kostic AD,et al. Leveraging vibration of effects analysis for robust discovery in observational biomedical data science. PLoS Biol 2021;19(9):e3001398.
7. Patel CJ, Burford B, Ioannidis JP. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J Clin Epidemiol 2015;68:1046e58.
8. Chu L, Ioannidis JPA, Egilman AC, Vasiliou V, Ross JS, Wallach JD. Vibration of effects in epidemiologic studies of alcohol consumption and breast cancer risk. Int J Epidemiol 2020;49:608e18.
9. Hoogeveen S, Sarafoglou A, Aczel B, Aditya Y, Alayan AJ, Allen PJ, et al. A many-analysts approach to the relation between religiosity and well-being. Religion Brain Behav 2022;13:1e47.
10. Kumar V, Yildirim A Ö, Nawroth P: The role of DNA damage in diabetic complications. Nat Rev Endocrinol. 2024 20(11):629-630. doi: 10.1038/s41574-024-01038-z.
11. Breznau N, Rinke EM, Wuttke A, Nguyen HHV, Adem M, Adriaans J, et al. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. Proc Natl Acad Sci U S A 2022;119:e2203150119.
12. Silberzahn R, Uhlmann EL, Martin DP, Anselmi P, Aust F, Awtrey E, et al. Many analysts, one data set: making transparent how variations in analytic choices affect results. Adv Methods Pract Psychol Sci2018;1(3):337e56.
13. Madigan D, Ryan PB, Schuemie M. Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies. Ther Adv Drug Saf 2013;4(2):53e62.
14. Zeraatkar D, Cheung K, Milio K, Zworth M, Gupta A, Bhasin A, et al. Methods for the selection of covariates in nutritional epidemiology studies: a meta-epidemiological review. Curr Dev Nutr 2019;3(10):nzz104.
15. van Dongen NNN, van Doorn JB, Gronau QF, van Ravenzwaaij D, Hoekstra R, Haucke MN, et al. Multiple perspectives on inference for two simple statistical scenarios. Am Stat 2019;73(sup1):328e39.
16. Landy JF, Jia ML, Ding IL, Viganola D, Tierney W, Dreber A, et al. Crowdsourcing hypothesis tests: making transparent how design choices shape research results. Psychol Bull 2020;146(5):451e79.
17. Schilling KG, Rheault F, Petit L, Hansen CB, Nath V, Yeh F-C, et al. Tractography dissection variability: what happens when 42 groups dissect 14 white matter bundles on the same dataset? Neuroimage 2021;243:118502.
18. Low J, Ross JS, Ritchie JD, Gross CP, Lehman R, Lin H, et al. Comparison of two independent systematic reviews of trials of recombinant human bone morphogenetic protein-2 (rhBMP-2): the Yale Open Data Access Medtronic Project. Syst Rev 2017;6(1):28.
19. Scientific Pandemic Influenza Group on Modelling. SPI-M-O: consensus statement on COVID-19, 8 October 2020. 2020. Available at: https://www.gov.uk/government/publications/spi-m-o-consensusstatement-on-covid-19-2-february-2022. Accessed March 2, 2024.
20. Willett WC, Stampfer M, Tobias DK. Re: adjustment for energy intake in nutritional research: a causal inference perspective. Am J Clin Nutr 2022;116:608e9.
21. Tomova GD, Arnold KF, Gilthorpe MS, Tennant PWG. Adjustment for energy intake in nutritional research: a causal inference perspective. Am J Clin Nutr 2022;115:189e98.
22. Jimmy Chun Yu Louie and Jahar Bhowmik: The (ab)use of food frequency questionnaire data in substitution modelling in nutritional epidemiology: a critique. European Journal of Clinical Nutrition; https://doi.org/10.1038/s41430-026-01712-7. 22. Feb. 2026
23. Hall KD. Challenges of human nutrition research. Science 2020;367:1298e300.
24. Ioannidis JPA. Unreformed nutritional epidemiology: a lamp post in the dark forest. Eur J Epidemiol 2019;34(4):327e31.
25. John P.A. Ioannidis Transparency, bias, and reproducibility across science: a metaresearch View. J Clin Invest. 2024;134(22):e181923. https://doi.org/10.1172/JCI181923
26. Ioannidis J.P.A.: The challenge of reforming nutritional epidemiologic research. Am. Med. Ass. Aug. 23, 2018; 10.1001/jama.2018.11025.
27. Ioannidis JPA, Berkwits M, Flanagin A, Bloom Th: Peer Review and Scientific Publication at a Crossroads. Call for Research for the 10th International Congress on Peer Review and Scientific Publication. JAMA September 22, 2023. doi:10.1001/jama.2023.17607
28. Chalmers I, Glasziou P: Avoidable waste in the production and reporting of research evidence. Lancet 2009; 374: 86–89. DOI:10.1016/S0140- 6736(09)60329-9
29. Garattini S, Chalmers I, Lind J: Patients and the public deserve big changes in evaluation of drugs: BMJ 2009;338:b1025; doi: https://doi.org/10.1136/bmj.b1025
30. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, Salman RA, Chan A-W, Glasziou P: Biomedical research: increasing value, reducing waste. Lancet 383: 101-104, 2014. http://dx.doi.org/10.1016/ S0140-6736(13)62329-6
31. Kleinert S. Horton R.: How should medical science change? Lancet 383: 197-198, 2014. http://dx.doi.org/10.1016/ S0140-6736(13)62678-1
32. Chalmers I, Bracken MB, Djulbegovic B, et al.: How to increase value and reduce waste when research priorities are set. Lancet 2014. http://dx.doi.org/10.1016/S0140-6736(13)62229-1.
33. Ioannidis JPA, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 2014. http://dx.doi.org/10.1016/S0140-6736(13)62227-8.
34. Salman R, Beller E, Kagan J, et al.: Increasing value and reducing waste in biomedical research regulation and management. Lancet 2014. http://dx.doi.org/10.1016/S0140-6736(13)62297-7.
35. Chan A-W, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. Lancet 2014. http://dx.doi.org/10.1016/S0140-6736(13)62296-5.
36. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet 2014. http://dx.doi.org/10.1016/S0140-6736(13)62228-X.
37. Schekman R. How journals like Nature, Cell, and Science are damaging science. The Guardian (London) Dec 9, 2013. http://www.theguardian.com/ commentisfree/2013/dec/09/how-journals-nature-science-cell-damagescience.
38. Aitkenhead D, Higgs P: I wouldn't be productive enough for today's academic system. The Guardian (London) Dec 6, 2013. http://www. theguardian.com/science/2013/dec/06/peter-higgs-boson-academicsystem.

nisfortschritt. Ärzte müssen daher bereit sein zu akzeptieren, dass Studien nie mit endgültiger Gewissheit beweisen können, dass ihre Ergebnisse mit der Wirklichkeit übereinstimmen. Die Möglichkeit des Irrtums (Fallibilismus) und die Heterogenität von Menschen, die in Studien multifaktorieller Erkrankungen nie vollständig erfasst werden kann, machen den Reiz des Arztberufes aus. Ein Reiz, der sich aus dem „cavere" bei Hippokrates und im „Grilling the data" erkennen lässt. <<

„Grilling the data": A call for a closer look at the results of clinical trials

Introduction: A doctor wishes to help his patient. The fundamental ethos of the medical profession, traced back to Hippocratic traditions—'primum non nocere'= not to do harm - is complemented by a second imperative: 'secundum cavere' = to be cautious. This is often understood merely as a call to examine the patient's problem closely. Another interpretation is: a commitment to a profound understanding of what one is doing. This can be achieved only through independent reading and interpretation of the original literature. To put it in modern terms: 'Critically explore, take responsibility through reflection'.

Methods: It is not enough to attend conferences for further training, to be familiar with guidelines, or to listen to the pharmaceutical representative's sales pitch. To do justice to the approach of 'critically exploring, taking responsibility through reflection' = 'cavere', one must be willing and able to understand the original works and to question the scientists' interpretations.

Results: This should be an important part of a doctor's work. It must not be left solely to lobbyists and professional societies, since the clinically practising doctor's independent, critical examination of the original data often leads to different conclusions than those of scientists, experts and professional societies.

If a significant number of doctors were to read at least some of the original studies, this would have an impact, experts and stakeholders could not ignore: criteria that are clinically relevant from the patient's perspective would replace those that, from a professional policy perspective, often lead scientists—enthused by their own research—to over-interpret their findings.

Discussion: An example is provided to illustrate the opportunity this presents for enhancing the standing of the medical profession and improving patient care.

Keywords

Education, critical rationalism, statistics, physician, patient care

Autorenerklärung

Der Autor erklärt, dass keine Interessenkonflikte vorliegen.

Zitationshinweis

Nawroth, P.: „Grilling the data: Ein Plädoyer für einen intensiveren Blick auf die Ergebnisse klinischer Studien", in: „Monitor Versorgungsforschung" (MVF 04/26), S. 82–86. http://doi.org/10.24945/MVF.04.26.1866-0533.2852

Prof. Dr. med. Dr. h.c. Peter Paul Nawroth ORCID: 0000-0002-6134-7804
leitete als Ärztlicher Direktor die Abteilung Innere Medizin I und Klinische Chemie am Universitätsklinikum Heidelberg und gehörte dem Klinikumsvorstand an. Seine Forschung gilt den Mechanismen chronischer Stoffwechselerkrankungen, mit besonderem Schwerpunkt auf diabetischen Folgeerkrankungen. Dazu leitete er einen Sonderforschungsbereich der Deutschen Forschungsgemeinschaft. Seit seiner Emeritierung ist er als wissenschaftlicher Autor tätig. Kontakt: PeterPaul.Nawroth@gmail.com

