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

‘Grilling the data’: A call for a closer look at the results of clinical trials

Summary

Introduction: A doctor wishes to help their patient. The fundamental ethos of the medical profession, traced back to Hippocratic traditions – ‘primum non nocere’ – 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 that it is an obligation to gain a profound understanding of what one is doing. This can only be achieved through independent reading and interpretation of the original literature. To put it in modern terms: ‘explore critically, 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 sales pitches of pharmaceutical representatives. To do justice to the approach of ‘critically exploring, taking responsibility through reflection’ = ‘cavere’, one must be willing and able to understand the original research and to question the scientists’ interpretations.

Results: This should be a key part of a doctor’s work. It must not be left solely to lobbyists and professional bodies, as the clinically practising doctor’s independent, critical examination of the original data often leads to different conclusions from those of scientists, experts and professional bodies. If many doctors were to read at least some of the original studies, this would have an impact that experts and stakeholders could not ignore: the criteria that are clinically relevant from the patient’s perspective would replace those which, from a professional policy perspective, often lead to over-interpretation by scientists who are enthusiastic about their own research.

Discussion: An example is used to illustrate the opportunity to enhance the standing of the medical profession and improve patient care.

Keywords

Continuing professional development, critical rationalism, statistics, doctor, patient care

Crossref/doi https://doi.org/10.24945/MVF.04.26.1866-0533.2852

Primum non (or nil) nocere – ‘first, do no harm’ – is a principle at the heart of ethical medical practice. Reading the full quote reveals further insights: ‘Primum non nocere, secundum cavere, tertium sanare.’ – “First, do no harm; second, be cautious; third, cure.” It is irrelevant that this quotation does not appear exactly as such in the writings of Hippocrates of Kos, and whether the Latin formulation originates from the Roman physician Scribonius Largus at the court of Emperor Tiberius or from the Englishman Thomas Sydenham. First and foremost, it is about a fundamental normative principle: under no circumstances should harm be caused. Then comes the approach: to look closely. To ensure that a doctor’s actions are more likely to benefit a person than to harm them. Only in third place comes the goal: to heal. For ‘secundum cavere’, merely ascertaining what is ailing the patient is not sufficient. The saying‘ Bona diagnosis, bona*curatio*’ is true, but it is too one-sided in the light of the possibilities offered by evidence-based medicine. A doctor practising caution can only fulfil this standard if they are prepared to read the original scientific data themselves, with the patient’s well-being in mind, and to question it critically. A doctor does not do justice to ‘secundum cavere’ if they trust, in good faith, the statements of the experts who deliver training courses and draft guidelines. The same scepticism must also be applied to lobbyists from professional bodies or corporate advertising. A doctor who meets the standard of prudence at this level too, bringing the necessary knowledge and a focus on the patient’s well-being to the discussion, achieves two things: Firstly, a deepening and increased credibility of the doctor-patient consultation; and secondly, a change in the approach taken by experts and guideline authors, speakers at conferences and training courses, and those who decide on marketing authorisation and reimbursement. Put in modern terms, this requirement for the doctor is: ‘secundum cavere’ = ‘to explore critically, to take responsibility through reflection’.

>> Is the individualisation of the normatively grounded duty of medical care – which goes far beyond assuming responsibility through adherence to guidelines and continuing professional development – feasible for a clinician? No one can, alongside their practical

The task involves reading all studies relevant to one’s practice and reflecting on their interpretation. Nevertheless, the impact will be dramatic if each of us were to read just one original paper a week and examine the data—including supplements—from the perspective of a doctor responsible for their patients: They will no longer be satisfied with what experts – often in collusion with companies and professional lobbyists – present to them. The result: the experts, companies and lobby groups will adapt to the new requirements, provided that the central demand remains: Study data should be examined with the patient’s well-being in mind. The following can be disregarded when assessing clinical efficacy relevant to patients: p-values that are irrelevant to this assessment, the mention of relative risk reduction (without specifying the absolute risk), arguments based on surrogate parameters – as these are not clinically relevant endpoints – and assumptions based on correlations – as these should only be presented as hypotheses. Efficacy can only be assessed if, for example, a study specifies the time to endpoint, calculates the effect size, and details non-responders, etc.

The following section uses ‘just’ one example (others are discussed elsewhere 1–2) to illustrate how surprisingly different – and how enriching for doctors, patients and the scientific community alike – a more open, patient-centred perspective on familiar data can be.

Methods and analytical approaches: 'Grilling the data'

Wang and colleagues published a paper entitled 'Grilling the data' (3), in which they analysed data from the National

Health and Nutrition Examination Survey (NHANES) from 2007 to 2014 in a different way to that usually employed in nutritional epidemiology (4–5). Using the 'specification curve analysis' approach —which is unfamiliar to us as doctors—they were able to circumvent the pitfall that is otherwise hidden from readers when interpreting observational studies: the various prior assumptions that scientists make in order to arrive at a specific result. These various assumptions, which influence the outcome of an analysis (3, 6–9), can be revealed using specification curve analysis. Consequently, one does not obtain a single, precise answer, but rather a separate result for each of the plausible assumptions. The advantage of this method is that, rather than a single point of pseudo-certainty, the reader is presented with a corridor illustrating the various possibilities underlying the data – a clearer picture than that provided by a confidence interval. The matter is, in fact, far more complex than scientists simplistically conclude: 'Factor Y increases the risk of a disease by X per cent, and many lives can be saved by reducing this harmful behaviour'.

Diseases with many risk predictors – such as cardiovascular disease, dementia or cancer (a correlation with a disease means we can only speak of predictors, not causal factors) – are difficult to study in observational studies. This is due to the many confounding factors and the way they influence one another. If people who eat more red meat weigh more and smoke more often, this does not mean that they smoke and weigh more because of the red meat. There are mathematical methods for accounting for these interdependencies, but the result remains merely a correlation, which depends on the model chosen. It is therefore not surprising that, for example, whilst intervention studies can reduce the incidence of diabetes by almost 50 per cent, they cannot reduce serious cardiovascular outcomes.

Furthermore, there are two problems with each of the risk predictors factored into the model: firstly, the results of the baseline studies show a much lower absolute risk reduction with semaglutide. That alone should be reason enough not to trust over-optimistic modelling and simulations! A further problem is that dietary habits are linked to education, social status, psychosocial and other factors, the interplay between which nobody understands with any precision and which, for methodological reasons, cannot all be analysed in a single study. This, in turn, leads back to the problem of uncontrolled confounding factors, which become particularly significant when the effect on clinically relevant endpoints (as is the case in most dietary and lifestyle studies) is only small. Finally, it must be borne in mind that the underlying data from observational studies may be falsified (more often than assumed), as reported in a recently published, as yet unreviewed study from Birmingham on strokes.

Furthermore, modelling based on large cohorts fails to account for the considerable heterogeneity among people: the differences in the metabolism of glucose, lipids, vitamins, etc. are far greater than the differences between fast food and the Mediterranean diet. This explains why a mathematically correct calculation does not necessarily mean that the result will stand up to clinical scrutiny.

When this is discussed with the authors of large cohort studies, the argument put forward is that one can only offer advice to policymakers if population-based studies are provided with simulations that are as accurate and comprehensive as possible. What about small but well-designed intervention studies to test the clinical efficacy of assumptions derived from large cohorts? Are doctors and policymakers really dependent on a simplification of study interpretations?

The result of a specification analysis, depicted as a broad corridor (Figure 1), provides a balanced view of the question of whether red meat is harmful or not, once all plausible prior assumptions have been disclosed:

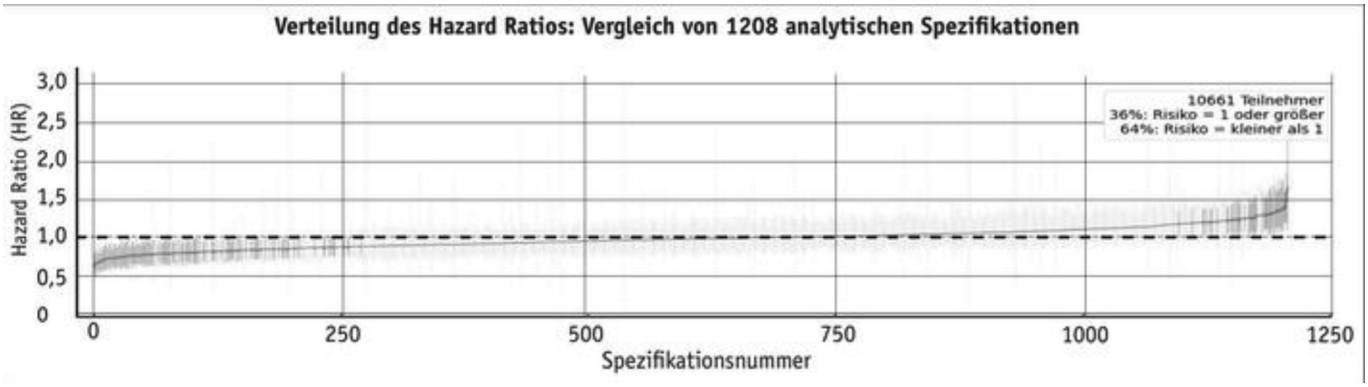


Fig. 1: Example of a wide corridor from a specification analysis. Key: The red data points lying above the line at 1 indicate an increased risk, for example from the consumption of red meat; those below 1 indicate a reduced risk. Fig. modified from (3) Wang et al. (2024) 'Grilling the data: application of specification curve analysis to red meat and all-cause mortality'. J Clin Epidemiol. Fig. 2.

The figure shows that, in many calculations , red meat has a fairly wide confidence interval; and secondly, it does not take into account that, for example, a 5 per cent increase in weight raises cardiovascular risk more than weight loss in intervention studies reduces the risk (see the difference between cardiovascular risk associated with an increase in body weight and that associated with weight loss; red meat may not be harmful, or may even be more beneficial than harmful. But beware: the majority of possibilities does not determine the truth!

Just because more calculations do not indicate an increased or reduced risk does not mean that there could be no danger associated with red meat. It would therefore be right for scientists to advocate for the funding of further studies to identify subgroups for whom red meat is either beneficial or harmful. This shows that scientists are hindering scientific progress by making unwarranted definitive statements instead of addressing open questions. Obvious avenues for advancing our understanding would include comparing those with a baseline elevation in LDL cholesterol or a glucose level at the upper end of the normal range, or, as the icing on the cake: How about investigating whether those whose LDL, glucose or markers of DNA damage (10) rise particularly sharply after consuming red meat compared to white meat are the ones who suffer more heart attacks, cancer and earlier death when consuming red meat? Such data would be more useful for a doctor-patient consultation than the scientifically oversimplified statement: 'Red meat could be harmful to you.'

This observation applies not only to red meat, but also to many other highly touted preventive measures. When one considers that even experienced scientists often disagree on which is the correct and best method of data analysis (3–28), it becomes clear

that we have become accustomed (perhaps due to a lack of critical questioning?) to oversimplified scientific statements. When one also considers that 62 per cent of dietary studies contain no validation of the quality of dietary history, and that deviations from reference methods can exceed 400 per cent (22), whilst the differences found often do not exceed the 10 per cent mark, it becomes clear that there must be serious doubts about the validity of many dietary studies (23–26). Couldn't presenting the broad range of possibilities, rather than sensationalist headlines, dispel the mistrust towards clinical recommendations based on epidemiological observations?

Discussion

Does the disclosure of uncertainty harm or benefit clinical practice?

Only at first glance does the presentation of data as a range appear to contradict the task of medical science to provide practical guidance to practising doctors. The 'secundum cavere' does not undermine the claim of the 'tertium sanare' for several reasons:

1. One should not expect more from a science than its methods and models allow. Otherwise, one endangers the patient, as the weighing up of risks based on an awareness of gaps in knowledge gives way to a false sense of security. From observational studies, one can only expect correlations that may or may not be based on causality. No matter how sophisticated and complex the calculations may be.

2. The broad range of possibilities gives the doctor the opportunity to make individual decisions and to work with their patient to seek tailored solutions in cases where – as is so often the case – there is no study data available that is suitable for that particular patient. This is what sets the doctor apart from 'Doctor Google'.

3. For the doctor, the 'cavere' means examining studies so closely that they do not trust interpretations of data at face value or even recklessly, and do not apply them uncritically to their patients.

4. By reading the literature independently, the practising doctor encourages the responsible scientists and experts not to give the false impression of certainty where, due to the methodology used, such certainty does not exist.

Conclusions:

Independent, critical reading – even if it is only of isolated studies – is advisable: who has not at some point suspected that there is a consensus amongst established scientists in various specialist fields, and thus also amongst peer reviewers, on how to interpret data, even though this one-sided form of analysis represents only a small section of the range of different, equally valid results (27–32)? What role do publishers play in this, given that they are keen to publish manuscripts that are widely cited and generate significant press coverage and advertising revenue? These weaknesses in the peer-review system as it is practised today are further exacerbated by the emergence of an ever-increasing number of sub-specialised journals, for which reviewers are far too often recommended by the relevant professional societies. It is not uncommon for the peer-review system to degenerate into a process of mutual evaluation by scientists who share the same views and similar interests (31–37). Everyone is aware of these weaknesses, but how do we adapt to them in such a way that we do justice to 'cavere' as well as to 'non nocere', without losing sight of 'sanare'?

The following answers apply:

1. By training ourselves to read and interpret original studies, so that we are not led up the garden path by biased experts who are part of the relevant academic circle. One aim: to quickly recognise why 85 per cent of the published studies in a Lancet series were described as 'rubbish' (30–36).

2. By reforming medical education (whilst resolving existing copyright issues so that original data can be shown to students): in addition to the statistics course, the joint reading of original data should become routine throughout all clinical semesters.

3. By insisting on patient-relevant criteria in publications (e.g. extent of endpoint delay, characterisation of non-responders, calculation of effect size rather than the p-value, specification curve analysis – where appropriate, subgroup analysis ...).

4. By insisting on the disclosure of the range of possible outcomes, we recognise not only the uncertainty of the findings but also the need for future research – a benefit for patients, researchers and the credibility of science.

5. By practising living with greater uncertainty (rather than the 'cookbook medicine' favoured by health insurance funds and guidelines), we enhance medical responsibility, herald the re-academisation of the medical profession and maintain the capacity for dialogue (e.g. with the IQWIG).

6. If European

universities (see the Stockholm Declaration), all journals would require the same formatting, tens of thousands of working hours would be saved for staff, and nonsensical publication incentives (37–38) would be abolished.

7. By insisting that clinically relevant studies are not primarily funded by companies, but that the scope for state funding of research is drastically expanded, so that the unresolved questions arising from studies submitted by companies can be investigated independently and using methodologically sound approaches, with patient-relevant endpoints. This research funding will cost less than the

vast sums spent on inappropriate medication. Evidence-based medicine, consistently applied and supported, reduces costs far more than any austerity package that fails to address the underlying issues. The task, therefore, is this: each of us should play our part in ensuring that the unresolved questions are brought to light, so that science is given the opportunity to shape a better tomorrow than today through new trials. In doing so, one prerequisite must be upheld: induction – that is, drawing generalisations from observations – is not a tool for advancing knowledge unless it is verified against other cases . Doctors must therefore be prepared to accept that studies can never prove with absolute certainty that their results correspond to reality. The possibility of error (failibilism) and the heterogeneity of human beings – which can never be fully captured in studies of multifactorial diseases – are what make the medical profession so compelling. An appeal that can be recognised in Hippocrates' 'cavere' and in 'grilling the data'. <<

References



‘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, traceable to Hippocratic traditions —‘primum non nocere’ = do no 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 bodies, as the independent, critical examination of the original data by practising doctors often leads to conclusions that differ from those of scientists, experts and professional bodies.

If a significant number of doctors were to read at least some of the original studies, this would have an impact that experts and stakeholders could not ignore: criteria that are clinically relevant from the patient’s perspective would replace those which, 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

Author’s declaration

The author declares that there are no conflicts of interest.

Citation

Nawroth, P.: ‘Grilling the data: A plea for a closer look at the results of clinical trials’, in: ‘Monitor Versorgungsforschung’ (MVF 04/26), pp. 82–86. https://doi.org/10.24945/MVF.04.26.1866-0533.2852

Prof. Dr med. Dr h.c. Peter Paul Nawroth ORCID: 0000-0002-6134-7804 served as Medical Director of the Department of Internal Medicine I and Clinical Chemistry at Heidelberg University Hospital and was a member of the hospital’s Executive Board. His research focuses on the mechanisms of chronic metabolic diseases, with a particular emphasis on diabetic complications. To this end, he headed a Collaborative Research Centre funded by the German Research Foundation (DFG). Since his retirement, he has been working as a scientific author. Contact: Peter-Paul.Nawroth@gmail.com

