



Discussion

Red herrings about relative measures: A response to Hoefler and Krauss

Jacob Stegenga



Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge, CB2 3RH, United Kingdom

In a discussion note published in this journal, [Hoefler and Krauss \(2021\)](#) criticise an article of mine published some years ago, also in this journal ([Stegenga 2015](#)). I welcome criticism, but their discussion note seriously misrepresents my work. Hoefler and Krauss neglect all of the fundamental arguments in the article they criticise, while wrongly accusing me of scholarly blunders. This is my rejoinder.

The issue under debate is about relative versus absolute measures of effectiveness, which are two families of ways of summarising data from empirical studies of interventions. In [Stegenga \(2015\)](#), [Kenna and Stegenga \(2017\)](#), and [Sprenger and Stegenga \(2017\)](#), we offered three distinct arguments to establish the epistemic and decision-theoretic superiority of absolute outcome measures compared with relative outcome measures. In particular, we argued that when a binary outcome is measured in a trial, the measure known as ‘risk difference’ or sometimes referred to as ‘absolute risk reduction’ is the best outcome measure. We further argued that absolute outcome measures should always be reported; I also argued that only sometimes should relative outcome measures be reported. Hoefler and Krauss dispute this latter point, asserting that absolute and relative measures should both always be reported.

Their discussion note begins with a criticism of my use of an empirical article which I had mentioned very briefly to illustrate the difference between absolute and relative measures ([Black et al., 1996](#)). Hoefler and Krauss dwell on irrelevant details of the Black article that divert attention away from my central arguments. They claim that I mischaracterised the Black article in my (2015), and they spend much more space than I did describing details of the Black article. Yet none of those details are relevant to the philosophical arguments. In the three articles cited above I devoted many pages to articulating technical arguments establishing the epistemic and decision-theoretic superiority of absolute outcome measures. In a few sentences in my 2015 article I cited data from Black to *illustrate* the formal difference between absolute and relative outcome measures. Data from the Black article in fact serves that illustration well, as Hoefler and Krauss themselves admit (p. 281). I could have simply invented a case to illustrate this—absolutely nothing about my philosophical arguments depended on any particular details of the Black article. Hoefler and

Krauss here focus on red herring irrelevant details about the Black article rather than on the fundamental question about absolute versus relative outcome measures.

Moreover, Hoefler and Krauss’s claim that I mischaracterised the Black article is itself a mischaracterisation. One of the reported outcomes of the Black article was about hip fractures. Table 3 of Black reports these data, and it was this information that I cited in my (2015). In short, Black said x , and I said that Black said x . Now Hoefler and Krauss claim that I mischaracterise Black. Their reasoning is as follows: Black reported data on multiple outcomes, but I only discussed data about hip fractures—Hoefler and Krauss conclude that “Stegenga’s presentation of only the latter constitutes a distortion of the study findings of [Black et al. \(1996\)](#).” However, the point of my discussion of Black was not to summarise the full range of findings in Black—if one wants to know what Black found, one can just read Black. The point, again, was merely to illustrate the difference between absolute and relative outcome measures, and I faithfully reported the data on hip fractures from Black for that purpose.

The broader context of the Black article renders their complaint even more injudicious. Hip fractures were the most important clinical outcome measured in the Fracture Intervention Trial that Black reported on ([Philipps 1997](#)). In the marketing campaign for alendronate, the Black article was widely appealed to as evidence that alendronate is effective at reducing the risk of hip fractures, and this data was reported and widely cited as a relative outcome measure. So I do not see the problem in my use of their data on hip fractures as illustration of the problems of relative outcome measures.

Early in their discussion note they give a paragraph-length quote of my (2015)—this was my only mention of alendronate in my entire article—in this paragraph, after contrasting the drug’s spuriously large relative effect size (50%) and its tiny absolute effect size (1%), I asserted that its tiny absolute effect size suggests that “alendronate sodium is barely effective” (p. 67). Hoefler and Krauss pick up on these five words, quote them repeatedly, and direct many paragraphs to challenging them. Those five words, taken in isolation, perhaps lend themselves to a misguided reading of this passage, which we could call *Drug Advice*: on this reading, one of my primary intentions was to make assertions about

E-mail address: jms303@cam.ac.uk.

<https://doi.org/10.1016/j.shpsa.2022.01.004>

Received 30 July 2021; Received in revised form 18 December 2021

the drug alendronate based on results of the Black article. In their discussion note Hoefler and Krauss appear to have the ambition of setting the record straight about both the Black article and the drug. They are, apparently, responding to the *Drug Advice* reading of those five words. Hoefler and Krauss worry that patients will be misled by my claims about alendronate, and physicians and medical scientists will view those five words as “ill-informed, factually mistaken, or misleading” (p. 281). Yet, *Drug Advice* is a misguided and uncharitable reading of those five words. Contrary to their worry at the end of their §1, patients do not routinely consult this journal for advice on drugs, and most medical scientists who I know—unlike Hoefler and Krauss, unfortunately—understand an illustrative claim in its wider context to stay focused on the substantive arguments.

There is a charitable reading of those five words, which I have already mentioned: that quoted paragraph was merely an illustration of the difference between absolute and relative measures—we could call this the *Argument Illustration* reading. My article was published in a philosophy journal, written by a philosopher of science, advancing philosophical arguments. It briefly used reported data about alendronate—among several other examples—merely to illustrate the subject. The paragraph in question starts with the words “to illustrate . . .”; the relevant section of my article from which this quote is taken is called ‘Measures’; all of the arguments in that section are about outcome measures; Hoefler and Krauss’s note is itself framed as being about outcome measures. The *Argument Illustration* reading should have been obvious. But Hoefler and Krauss treat this section of my paper as a medical paper about alendronate, and complain that it does not give a full clinical picture of the benefits of the drug, when such a picture is entirely irrelevant to the philosophical point. If I had intended *Drug Advice*—if I had wanted to make an argument about alendronate—I would have discussed a wealth of other relevant considerations, such as empirical work more recent than a twenty-year old trial, the recent studies investigating the drug’s harms, the thousands of lawsuits raised against the drug, the views of the experts who hold that standards for diagnosis with osteoporosis have been excessively weakened, and the views of other experts who promote the merits of non-pharmaceutical means for avoiding hip fractures. I did not discuss any of that because none of it was relevant to the arguments of my (2015).

Yet, Hoefler and Krauss seem to be committed to the misguided *Drug Advice* reading of this passage when they conclude: “Stegenga’s discussion of alendronate is an example of the kind of mistakes that philosophers of science need to avoid committing” (281). They seem to be ascribing to me the inference from a particularly small absolute effect size to a conclusion about relatively low effectiveness. Those five words Hoefler and Krauss dwell on do lend themselves to that ascription. Yet, part of the point of my (2015) was to argue that we *cannot* directly make inferences about effectiveness based on effect sizes alone. No one who read that article (or subsequent relevant work) should have been tempted to think that I was claiming we can slide from effect sizes to claims about effectiveness. Under the obvious *Argument Illustration* reading of those five words, I made no mistake. Hoefler and Krauss are talking directly past the core content of my (2015) and subsequent work—it is their commitment to *Drug Advice* at the expense of *Argument Illustration* which is the real mistake.

Their mistakes are layered. In a footnote Hoefler and Krauss speculate about “the source of Stegenga’s idea that Black et al. (1996) studied hip fractures.” This implies that Black did not study hip fractures and I ignorantly thought that Black did study hip fractures. Yet, Black did study hip fractures! Moreover, the Black publication itself is packaged as highly relevant to hip fractures; hip fractures are mentioned in the abstract, introduction, the discussion section, and supporting data. Discussion of hip fractures appears literally on every page of the Black article, in which the authors write things like “Hip fractures have the most serious consequences and account for most of the costs of osteoporosis; our findings suggest that use of alendronate for women at increased risk of hip fracture might reduce the disability and the costs of osteoporotic fractures”

(1539). Hoefler and Krauss have falsely ascribed an error to me and then speculated about the source of that putative error, while layering their own mistakes on top of each other.

The arguments in Stegenga (2015), Kenna and Stegenga (2017), and Sprenger and Stegenga (2017) establish that absolute outcome measures are epistemically and decision-theoretically superior to relative measures. Hoefler and Krauss ignore all of the relevant arguments. Particularly egregious is their neglect of the decision-theoretic argument. That argument establishes the following: supposing the usual assumptions of decision theory are granted, if we wish to determine which of two treatments are best, then relative outcome measures are irrelevant, while absolute outcome measures are—on their own—sufficient. Perhaps Hoefler and Krauss disagree with principles of utility maximisation; perhaps they think that some medical contexts involve high enough risks such that risk-aversion renders the standard decision-theoretic approach problematic (and that somehow this would favor relative measures); perhaps there was another premise of the argument they found wanting. In any case, they did not discuss this argument or its premises at all. Hoefler and Krauss merely asserted that they disagree with its conclusion.

For the record, here are the other two arguments. First: absolute measures incorporate the background or prior probability of the outcome of interest, but relative measures do not, so if one wants to make an inference about the probability of the outcome both with and without treatment, absolute measures are sufficient, while relative measures lend themselves to a fallacy much like the base-rate fallacy (insofar as the prior probability of the outcome is neglected). Second: we often wish to estimate the ‘causal strength’ of interventions, and absolute measures have attractive and intuitively desirable properties in that regard while relative measures lack those properties (see Sprenger, 2018 for a full statement of this argument). Like the decision-theoretic argument, Hoefler and Krauss do not address these arguments.

Though of course I stand by those three arguments, it would be hubris to think that is the end of the matter about outcome measures. There are many ways in which the subject could be advanced. For example, Ina Jäntgen has argued (in an as-yet unpublished note) that the decision-theoretic superiority of absolute outcome measures over relative outcome measures is more constrained than originally suggested by Sprenger and Stegenga (2017). Jäntgen argues that when comparing two interventions which are tested in different trials, there are conditions under which both absolute and relative measures are sufficient for deciding between the interventions. Such impressive work may indeed demonstrate limitations on my earlier conclusions, a result I would welcome with open arms.

Let’s go back to Black. Hoefler and Krauss assert that my “central claim and recommendation about the importance of absolute over relative outcome measures” was “derived from a misleading analysis of alendronate research” (283). At this point, I hope that this assertion surprises others as much as it surprises me. My three principled arguments have disappeared, and I am saddled with deriving a general methodological prescription from a single illustrative case study (one which I spent a mere handful of lines on).

Hoefler and Krauss focus on what outcome measures should be reported, and they do not engage with my prior arguments for the epistemic or decision-theoretic superiority of absolute outcome measures. Yet, nearly all of my arguments and by far the vast majority of devoted space in the works in question were directed toward the superiority of absolute outcome measures rather than their reporting. This is yet another instance in which Hoefler and Krauss are talking past the content of my (2015) and subsequent articles.

What should we say about the reporting of outcome measures? To move from “absolute outcome measures are epistemically and decision-theoretically superior to relative outcome measures” to “absolute outcome measures should always be reported and relative outcome measures should only sometimes be reported” requires further premises. One such premise which I drew on is the tendency of physicians and patients to overestimate the effectiveness of medical interventions when

presented with relative outcome measures, which is an empirically substantiated phenomenon. Another premise could involve appealing to ethical or political considerations which could inform how scientists should present their results. A forthcoming article by Schroeder (2022) does exactly this, and indeed Schroeder concludes that, given a very modest ethical principle (namely, that physicians and scientists should promote informed decision-making), the arguments in Sprenger and Stegenga (2017) establish that absolute outcome measures should be reported. Though Schroeder does not explicitly add the constraint regarding relative outcome measures, he does argue that irrelevant information should not be reported, and the arguments in Sprenger and Stegenga (2017) conclude that relative outcome measures are irrelevant to decision-makers.

Hoefer and Krauss claim that the recommendation to report absolute and not relative measures amounts to “concealing one sort of information in favor of another” (282)—similarly, they assert that to not report relative measures is to “withhold data”. To be clear, a relative measure is not “data”, it is a quantitative summary of data, and thus not reporting a relative measure certainly does not amount to withholding data. There are an infinite number of possible outcome measures, most of which are never reported in any trials, yet presumably Hoefer and Krauss would not want to claim that all trials are withholding an infinite amount of data. More to the point, their claim here straightforwardly begs the question against the arguments that they failed to engage with. From the perspective of an individual decision-maker, assuming known patient utilities, all of the information which is necessary and sufficient to make an informed treatment decision is contained in the absolute measure, and this is not true for the relative measure; conversely, relative measures are misleading for individual decision-makers. Hoefer and Krauss have not said what sort of information they think relative measures convey; on the other hand, the arguments that they neglect conclude that relative measures do not in fact convey relevant information when choosing an intervention.

The argument in Hoefer and Krauss seems to be that sometimes we can get people to make better decisions by reporting relative outcome measures (282). It is worth emphasising how irrelevant this is to my thesis that absolute outcome measures are epistemically and decision-theoretically superior to relative outcome measures. We can get people to make better decisions in many ways—by deceiving, coercing, or bullying, for example. The Nazis campaigned against smoking as part of their pursuit of ‘racial hygiene’ (Proctor, 2000). Hoefer and Krauss seem to hold that the possibility of getting people to make decisions that are better for their health by reporting relative outcome measures is an argument for reporting relative outcome measures. Presumably they would not extend that argument to outright deception, though their reasoning applies just as much to deception as it does to reporting relative outcome measures.

Their argument about communicating relative measures to patients faces a serious problem, pointed out to me by Aaron Kenna. Hoefer and Krauss construct a case in which a person can lower their risk of a low probability/high disutility outcome with a vaccine that has “no, or limited, known negative side-effects” (282), and then Hoefer and Krauss claim that reporting the relative risk reduction (80%) is more likely to get the person to use the vaccine than if only the absolute risk reduction (1.2%) were reported (and thus, we should report the relative measure). This of course depends on the person's risk profile, their utility assignments to the possible outcomes and the costs associated with taking the vaccine, their fear of needles, etc. Hoefer and Krauss can brush these concerns away by stipulating that the vaccine is entirely free, painless, not anxiety-inducing, and so on. But this amounts to constructing a hypothetical situation in which taking the vaccine is clearly the rational thing to do—all benefit and no cost! Here is the difficulty for Hoefer and Krauss: in such a situation any plausible outcome measure will show that taking the vaccine is rational. In particular, the physician need only communicate the absolute risk reduction, which alone is sufficient to show that taking the vaccine is the best thing to do. In the case they

construct, the debate about the merits of communicating relative versus absolute measures is moot. On the other hand, if the vaccine is not cost-free, as is any real medical intervention, then the physician should communicate information which allows the patient to determine the costs and benefits of taking versus not taking the vaccine. The only outcome measures which convey this type of information are, of course, absolute outcome measures.

What Hoefer and Krauss suggest is that, since not all patients “have a basic grasp of probability and statistics” (282), we ought to nudge patients to make better decisions by communicating relative outcome measures. Better according to what standard? Any plausible model of healthcare decision-making involves the physician and patient appealing to empirical facts and patient preferences, but not the preferences of physicians or philosophers. The consideration on offer by Hoefer and Krauss involves infringing on the ability of the patient to make informed, autonomous decisions.

Hoefer and Krauss claim that their prescription is consistent with the authoritative medical journal *The New England Journal of Medicine*, because *The New England Journal of Medicine* guidelines claim that both absolute and relative measures should be reported. They also seem to think that publications in *The New England Journal of Medicine* in fact report both outcome measures (282). They later claim that “Philosophy of science should engage with actual scientific practices” (283), which is ironic, because unfortunately *The New England Journal of Medicine* guidelines are not representative of actual scientific practices—not even in *The New England Journal of Medicine*. A recent analysis of the reporting of absolute and relative outcome measures in the world's six most-cited medical journals showed that trials with binary outcomes reported absolute measures only about one third of the time, and in *The New England Journal of Medicine* this proportion was substantially lower (Elliott et al., 2021). This contradicts their claim that “both absolute numbers and relative numbers tend to be reported” (281). The recommendation offered by Hoefer and Krauss is inconsistent with, and apparently out of touch with, scientific practice.

In fact, the prescription that both absolute and relative measures be reported is far from consensus among medical and scientific authorities. Here, for example, is the Association of the British Pharmaceutical Industry (ABPI) code of practice: “Referring only to relative risk, especially with regard to risk reduction, can make a medicine appear more effective than it actually is. In order to assess the clinical impact of an outcome, the reader also needs to know the absolute risk involved. In that regard relative risk should never be referred to without also referring to the absolute risk. Absolute risk can be referred to in isolation.” The UK Patient Information Forum, guided by the eminent statistician David Spiegelhalter, claims that patient communication should “use absolute rather than relative risk” (Patient Information Forum, 2019). Moreover, some authoritative guidance which appears to support the prescription of Hoefer and Krauss offers thin justification for it, claiming merely that both families of measures are informative (see, e.g., the CONSORT statement, §17b, on reporting binary outcomes).

Their discussion note goes beyond the usual norms of scholarly critique when they claim that I have done “a disservice to medical research and the philosophy of medicine” (280). This is the first of many instances throughout their discussion in which Hoefer and Krauss veer from scholarly commentary to insulting sermon. A truly puzzling criticism in their discussion amounts to listing things that I did not discuss: “Stegenga does not, in his (2015a) paper, mention the measures that have been taken to improve medical research, such as the requirements of pre-registration of trials and pre-publication of protocols, requirements of making trial outcome data publicly available even if negative, registered reports in which a study protocol is peer reviewed and provisionally guaranteed publication before the study is conducted, regardless of the results (to counteract publication bias), and so on” (283). The manner of this claim suggests that these tactics have been unequivocally successful in improving medical research, which, unfortunately, is far from true (as any work which genuinely paid attention to the practice of medical

research would be sensitive to). In any case, why these issues should have been discussed in the particular article Hoefer and Krauss are criticising is unclear. What do these issues have to do with outcome measures? Nevertheless, I have spent a good part of my career discussing these and related issues (see, e.g., Stegenga et al., 2017; Hitzig & Stegenga, 2020; Moynihan et al., 2019).

In their final paragraph Hoefer and Krauss claim that my proposal “is unlikely to be taken seriously by the greater medical community” (283). This speculation adds no intellectual content to the question of whether relative outcome measures should be reported. Anyway, there are all sorts of reasons why the medical community might not take seriously a philosophical proposal. It is similarly unlikely that the general physics community would read, let alone take seriously, a philosophical article on absolute and relational theories of space and motion, yet that fact is hardly relevant to assessing the merits of such an article.

Hoefer and Krauss immediately follow that irrelevant quip with another: “Likewise, Stegenga’s (2018) general thesis in *Medical Nihilism*, namely that we should have little confidence in the effectiveness of most medical treatments, goes against the simple historical fact that medical researchers have developed effective treatments for a wide range of diseases. Medical research has enabled us to cure illnesses, develop vaccines to prevent others and has contributed to expanding our life expectancy” (283). This passing snub is irrelevant to my position on outcome measures articulated in my 2015 article. Moreover, to dismiss any book with a wave of a hand is yet more poor form. To suggest that a book of this nature would not have considered these extremely basic points is surprising: the few cures that medicine has developed form the core of the ‘magic bullet’ concept I develop, which itself is part of the overall argument of the book; the claim that medicine has significantly improved life expectancy is contradicted by the famous and very well substantiated ‘McKeown Thesis’, which also contributes to the argument of the book; the “simple historical fact” they mention is a drop in the relevant inferential bucket.

The advertised thesis of their discussion note is about outcome measures. But too often in their discussion Hoefer and Krauss urge the secondary sermon that “philosophers should take greater care” (283). It looks so patronising when others write it; it looks so good when one

writes it oneself. Yet, I personally cannot bring myself to take those words out of quote and put them into print. Rather, I urge Hoefer and Krauss and any interested readers to consult the relevant arguments about absolute versus relative outcome measures.

Acknowledgements

I am grateful to Anna Alexandrova, Aaron Kenna, Boaz Miller, Andrew Schroeder, and two anonymous referees for comments.

References

- Black, D. M., Cummings, S. R., Karpf, D. B., Cauley, J. A., Thompson, D. E., Nevitt, M. C., et al. (1996). Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. Fracture Intervention Trial Research Group. *Lancet*, 348(9041), 1535–1541.
- Elliot, M., Skydel, J., & others. (2021). Characteristics and reporting of number needed to treat, number needed to harm, and absolute risk reduction in controlled clinical trials, 2001–2019. *JAMA Internal Medicine*, 181(2), 282–284.
- Hitzig, Z., & Stegenga, J. (2020). The problem of new evidence: P-hacking and pre-analysis plans. *Diametros*, 17(66), 10–33.
- Hoefer, C., & Krauss, A. (2021). Measures of effectiveness in medical research: Reporting both absolute and relative measures. *Studies in History and Philosophy of Science*, 88, 280–283.
- Kenna, A., & Stegenga, J. (2017). Absolute measures of effectiveness. In McClimans (Ed.), *Measurement in medicine: Philosophical Essays on Assessment and evaluation*. Rowman and Littlefield, 2017.
- Moynihan, R., Bero, L., & others. (2019). Pathways to independence: A plan to produce and use trustworthy evidence. *BMJ*, 367, l6576.
- Patient Information Forum. (2019). How to communicate benefits, risks, and uncertainties. <https://pifonline.org.uk/download/file/87/>.
- Phillips, A. (1997). The fracture intervention trial. *The Lancet*, 349(9050), P505.
- Proctor, R. (2000). *The nazi war on cancer*. Princeton University Press.
- Schroeder, A. (2022). An ethical framework for presenting scientific results to policy-makers. *Kennedy Institute of Ethics Journal*, 32(1).
- Sprenger, J. (2018). Foundations of a probabilistic theory of causal strength. *Philosophical Review*, 127(3), 371–398.
- Sprenger, J., & Stegenga, J. (2017). Three arguments for absolute outcome measures. *Philosophy of Science*, 84(5), 840–852.
- Stegenga, J. (2015). Measuring effectiveness. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 54, 62–71.
- Stegenga, J. (2018). *Medical nihilism*. Oxford: Oxford University Press.
- Stegenga, J., Kennedy, A. G., Tekin, S., Jukola, S., & Bluhm, R. (2017). New directions in philosophy of medicine. In Marcum (Ed.), *Bloomsbury companion to contemporary philosophy of medicine* (pp. 343–367). Bloomsbury.