Book Forum

Medical Nihilism by Jacob Stegenga: What is the right dose?

"We should have little confidence in the effectiveness of medical interventions" (2018, p. 167). This rattling statement is the thesis that Jacob Stegenga calls *medical nihilism* in his book of the same name. Stegenga's monograph is a philosophical analysis and criticism of contemporary medical research that dissects the concepts, research methods and social context of modern medicine. In building his case, Stegenga exposes vulnerabilities in our concepts and methods that have served as sites for various biases and interests to corrupt the evidence base for medical therapies, especially drugs.

Stegenga's book is analytic, persuasive, eminently readable and engages enthusiastically with the medical literature. It is a towering contribution to philosophy of science and the burgeoning field of philosophy of medicine. I trained in medicine as well as in philosophy. This book speaks forcefully to both parts of me. Serious criticism of medicine has often been left to other humanities disciplines. Important sociological, anthropological and historical critiques of medicine have been written, but no philosopher of science has wielded the tools of analytic philosophy to launch a book-length critique quite like Stegenga does in *Medical Nihilism*, which now belongs to a body of critical literature that includes classics like Illich's (1975) *Medical Nemesis*, Foucault's (1965) *Madness and Civilization*, Wootton's (2007) *Bad Medicine*, and McKeown's (1980) *The Role of Medicine: Dream, Mirage, or Nemesis?*

Compared to Illich, who was nihilistic about the entire medical enterprise ("The medical establishment has become a major threat to health" (Illich, 1975, p. 3)), Stegenga's 'medical nihilism' is a therapeutic nihilism, challenging the effectiveness of contemporary medical interventions. Should we accept his conclusion? Not quite, I argue. The problem is that there is vagueness or imprecision in his central argument, which stands in the way of meaningfully interpreting and evaluating it. Moreover, the existing meta-research evidence on medicine that might help make the terms in his argument more precise is not constraining enough to render it interpretable and evaluable. Thus, we cannot conclude that our confidence in the effectiveness of medical interventions – even contemporary pharmaceuticals – ought to be 'low'. Nonetheless, Stegenga's book succeeds in arguing that we should be *less* confident in the effectiveness of drugs than many of us are, which is a significant achievement, and demands response both from philosophy of science and from medicine.

Stegenga's claim is not the one that we should have low confidence in *all* medical treatments: some interventions like insulin and antibiotics are 'magic bullet's that work very well. Rather, we should have low confidence in general, in most interventions. One could rightly question whether Stegenga's nihilistic conclusion is the most helpful one to draw given that it does not tell us how to more reliably assess our confidence in interventions or indeed which interventions are magic and which are muck (but as we will see in a moment, Stegenga's central *arguments* do address these issues). However, here I scrutinize whether his main conclusion is warranted by his arguments.

One could question the wide scope of Stegenga's therapeutic nihilism. As a medical student, I was struck by the dizzying diversity of medical interventions, from surgeries and vaccines to psychotherapy and plain old medical counseling and advice. There is a diversity of kinds of evidence to go along with this diversity of treatments, including evidence from clinical experience, basic science and observational epidemiology (these kinds of evidence have their own problems). As Broadbent (2019) points out, Stegenga's criticisms of medical research best apply to pharmaceuticals studied in clinical trials. I agree, but here I challenge Stegenga's conclusion even if we qualify it as a narrower thesis about contemporary pharmaceuticals.

Stegenga uses Bayes' Rule to organize his Master Argument for medical nihilism:

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E)}$$

H is a hypothesis claiming that a medical intervention is effective. The evidence E is best understood as the apparent finding of a small effect size in a clinical trial or meta-analysis of trials. The conclusion of the Master Argument is that the probability of H given E, $p(H|E)$, is generally low: "on average we ought to have a low posterior probability in that hypothesis, $p(H|E)$ – in short, medical nihilism is compelling" (Stegenga, 2018, p. 168). The chapter-level arguments of Stegenga's book are meant to provide justification for thinking that the prior probability of H (or $p(H)$) is 'low', the probability of E (or $p(E)$) is 'high', and the probability of E given H (or $p(E|H)$) is 'low'; and therefore, by Bayes' Rule, $p(H|E)$ is 'low'.

$p(H)$ is low because the pathophysiologic basis of most diseases is complex (we should expect few interventions to be magic bullets that fly right to the heart of the disease and target it successfully with few side effects) and historically many medical interventions in use at some point have later been rejected as ineffective. $p(E)$ is high because of widespread research bias: clinical trials and meta-analysis are malleable and vested interests tune the methods to get a positive finding; thus, the probability of finding a small positive result is high. Finally, $p(E|H)$ is low because most apparent effect sizes measured in trials are low (low enough that we should actually attribute them to bias rather than any effectiveness of the intervention) and because studies are frequently discordant, contradicting each other – if most interventions were truly effective, these findings would be uncommon.

There is vagueness or imprecision in Stegenga's conclusion that our confidence in the effectiveness of medical interventions should be 'low'. What is meant by 'low'? My quantitation of 'low' might overlap substantially with your quantitation of 'high' (a fact that has led the US Intelligence Community to recommend the use of numbers in stating the probability of politically relevant hypotheses (Tetlock & Gardner, 2015)). This problem makes the implications of Stegenga's conclusion difficult to discern. Stegenga argues that "there can be no precise or general answer" to how low is 'low', but: "It is enough to say: lower, often much lower, than our confidence on average now appears to be", and he argues that our present confidence is high (p. 183). While the suggestion that our present confidence is inflated is important, as

Stegenga suggests it does not put a precise number to his medical nihilism. Sometimes, a low confidence is enough to justify using a drug, depending on what is to be gained or lost – it depends on how low is 'low'.

However, I already said that I would not focus on how helpful his thesis is, but rather whether it is warranted. The problem of imprecision I want to explore further is vagueness in the other terms of Stegenga's argument: p(H), p(E) and p(E|H). Stegenga writes, "I do not pretend that the terms in the master argument can be precisely quantified, either generally or for a particular medical intervention" (p. 178). Yet imprecision makes it challenging to evaluate Stegenga's arguments for each term. The fact that many medical interventions are eventually rejected and that the pathophysiology of many diseases is complex does not allow us to conclude that p(H) is low until we have a clear idea of what 'low' means (and how common these phenomena are – the pathophysiologic rationale for biologics such as monoclonal antibodies may be stronger than the rationale for antidepressants or deep brain stimulation). Even then it is challenging to draw a definite conclusion.

Moreover, imprecision makes it difficult to evaluate the overall Master Argument. Stegenga notes that the Master Argument is valid because it makes use of Bayes' Rule. But Bayes Rule is an equation, and the conclusion of the Master Argument only follows validly if we choose certain numerical values for 'low' and 'high'. To illustrate, if by 'low' probability we mean 0.5 and by 'high' probability we mean 0.25, then plugging in these numbers into Bayes' Rule we get a p(H|E) of 1.0, which does not meet our definition of 'low'. Though it might seem strange to consider 0.25 to be high and 0.5 to be low, 'high' and 'low' are being applied to different types of phenomena, so it might be reasonable to say that a 25% rate of bias among clinical studies is high while a 50% prior confidence in the hypothesis is low.

Adding precision to the terms of the Master Argument would make it more interpretable and evaluable. But then a tricky question presents: what numerical values for the terms of the Master Argument are warranted? Stegenga often discusses relevant evidence of wide trends in medical research, including industry bias, publication bias and discordance among studies. As this kind of research is relevant to assessing effectiveness, I have called it 'meta-research evidence' and argued that it should be used in evaluating therapies (Fuller, 2018). Meta-research evidence provides a promising avenue for providing quantitative constraints on the terms of the Master Argument and helping us better assess just how confident in the effectiveness of medical interventions we should be.[1]

Stegenga argues that the historical record of medical interventions that were once accepted and later rejected as ineffective provides support for a low prior probability of effectiveness, but adds that this phenomenon lacks a quantitative analysis. Prasad et al. (2013) analyzed all studies of medical practices published in the *New England Journal of Medicine* in the first decade of the Twenty-First Century. Of 363 tests of an existing standard medical practice, 146 (40%) found it worse than a previous standard or than doing nothing ('medical reversal'), 138 (38%) confirmed the standard and 79 (22%) were inconclusive. In other words, half of all conclusive tests were reversals, which might suggest an estimate for the frequency of untested interventions that are effective of 0.5 – the prior probability of effectiveness before testing. However, this 0.5 may not provide a reasonable estimate for p(H) for many reasons, including: the medical practices analyzed by Prasad et al. were not exclusively treatments, tests are not always correct, and the sample of practices studied might not be representative of all medical practices (excluding, for instance, newer or experimental interventions).

Turning to the probability of the evidence given the hypothesis, p(E|H), Stegenga notes that clinical trials and meta-analyses of the same intervention often contradict one another. He cites meta-research by Ioannidis (2005) showing that 14/34 (41%) of positive highly cited therapeutic studies that were compared with a second study of the same intervention were contradicted by the second study in terms of the qualitative (20.5%) or quantitative (20.5%) treatment effect. There are many meta-research studies of contradictions in the medical literature, some of them involving large numbers of interventions in the Cochrane database (e.g. Furukawa, Guyatt, & Griffith, 2002; Pereira, Horwitz, & Ioannidis, 2012; Pereira & Ioannidis, 2011). However, none of these studies pretend to include a representative sample of interventions, and there are many possible explanations when two studies disagree, not all of which suggest problems with the evidence (Fuller, 2018).

A further consideration Stegenga raises for determining the p(E|H) is that many apparent therapeutic effect sizes are small. He cites some examples of effect sizes listed on NNT.com. The number-needed-to-treat (NNT) is a measure of effect size, the number of people one would need to treat with the intervention to prevent or cause one outcome. I did a bit of meta-research of my own, and found that of the 70 unique interventions on NNT.com (most of them pharmaceuticals) deemed to be effective with listed NNTs, 35 (50%) had an NNT greater than 10, and 53 (76%) had an NNT greater than 5 (The NNT Group, 2019). Unfortunately, it is difficult to discern the relevance of these data because it is unclear how small an apparent effect size must be if we are to plausibly attribute it to bias.[2]

Finally, in quantifying the rate of bias among therapeutic studies and the probability of the evidence, p(E), several kinds of meta-research that Stegenga discusses are relevant. For instance, a Cochrane review by Lundh, Lexchin, Mintzes, Schroll, and Bero (2017) suggests that industry-funded drug and device studies are 1.27 times more likely to be positive compared to non-industry-funded studies. Much of Stegenga's analysis of medical research provides a convincing reason to believe that this association is due to bias ('industry bias'). Thus, we could estimate that the results of at least 27 of every 127 (21%) positive industry-funded trials are due to bias - that's 14% of all industry-funded trials included in the review. Other meta-research evidence such as evidence of widespread publication bias (Hopewell, Loudon, Clarke, Oxman, & Dickersin, 2009) suggests that the probability of bias might be higher than 14% but does not narrow the probability of bias on a more precise value. Moreover, the probability of positive evidence E depends on more than just the probability of bias.

In summary, the available meta-research evidence is not sufficient for estimating numeric values for the terms of Stegenga's Master Argument. On the other hand, the values he assigns to them ('low', 'high') are too imprecise to evaluate whether they are warranted and whether they in turn warrant the conclusion, 'medical nihilism'. However, Stegenga's arguments do succeed in showing that we should be *less* confident in the effectiveness of drugs than many of us are.

More meta-research could be useful for further constraining these values and evaluating Stegenga's therapeutic nihilism. However, there is a more urgent reason for wanting this kind of evidence: for the purpose of re-evaluating individual therapies (Fuller, 2018), especially considering the problems that Stegenga explores masterfully in his book. Stegenga's rigorous defense of his position in *Medical Nihilism* should temper any naïve medical optimism. To that end, it is an important antidote, even if the precise dose is difficult to titrate.

Acknowledgments

---

[1] Elsewhere, Stegenga (2011) provides a strong critique of meta-analysis, a tool often used in meta-research.

[2] In addition to 82 interventions listed as effective, NNT.com lists 47 interventions as ineffective on net, 16 interventions as harmful on net, and 39 interventions as lacking sufficient evidence to make a determination.

References

Broadbent, A. (2019). *Philosophy of medicine.* New York: Oxford University Press.

Foucault, M. (1965). *Madness and civilization: A history of insanity in the age of reason.* New York: Random House, Inc.

Fuller, J. (2018). Meta-research evidence for evaluating therapies. *Philosophy of Science, 85*, 767–780.

Furukawa, T. A., Guyatt, G. H., & Griffith, L. E. (2002). Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *International Journal of Epidemiology, 31*(1), 72–76.

Hopewell, S., Loudon, K., Clarke, M. J., Oxman, A. D., & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews, 1*, Mr000006. https://doi.org/10.1002/14651858.MR000006.pub3.

Illich, I. (1975). *Medical Nemesis: The expropriation of health.* London: Calder & Boyars Ltd.

Ioannidis, J. P. A. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association, 294*(2), 218–228.

Lundh, A., Lexchin, J., Mintzes, B., Schroll, J. B., & Bero, L. (2017). Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews, 2*, Mr000033. https://doi.org/10.1002/14651858.MR000033.pub3.

McKeown, T. (1980). *The Role of medicine: Dream, mirage, or Nemesis?* Princeton: Princeton University Press.

Pereira, T. V., Horwitz, R. I., & Ioannidis, J. P. (2012). Empirical evaluation of very large treatment effects of medical interventions. *Jama, 308*(16), 1676–1684.

Pereira, T. V., & Ioannidis, J. P. (2011). Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *Journal of Clinical Epidemiology, 64*(10), 1060–1069.

Prasad, V., Vandross, A., Toomey, C., Cheung, M., Rho, J., Quinn, S., ... Cifu, A. (2013). A decade of reversal: An analysis of 146 contradicted medical practices. *Mayo Clinic Proceedings, 88*(8), 790–798.

Stegenga, J. (2011). Is meta-analysis the platinum standard of evidence? *Studies in History and Philosophy of Biological and Biomedical Sciences, 42*(4), 497–507.

Stegenga, J. (2018). *Medical nihilism.* Oxford: Oxford University Press.

Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction.* New York: Broadway Books.

The NNT Group (2019). *Therapy (NNT) reviews by rating.* http://www.thennt.com/home-nnt/, Accessed date: 5 May 2019.

Wootton, D. (2007). *Bad medicine: Doctors doing harm since hippocrates.* Oxford: Oxford University Press.

Jonathan Fuller,
Department of History and Philosophy of Science, University of Pittsburgh,
Pittsburgh, PA, USA
E-mail address: JPF53@pitt.edu.