Meta-Research Evidence for Evaluating Therapies

Jonathan Fuller*†

The new field of meta-research investigates industry bias, publication bias, contradictions between studies, and other trends in medical research. I argue that its findings should be used as meta-evidence for evaluating therapies. 'Meta-evidence' is evidence about the support that direct 'first-order evidence' provides the hypothesis. I consider three objections to my proposal: the irrelevance objection, the screening-off objection, and the underdetermination objection. I argue that meta-research evidence works by rationally revising our confidence in first-order evidence and, consequently, in the hypothesis—typically, downward.

1. Meta-Research on the Problems with Medical Evidence. Problems in medical research abound, threatening to undermine our confidence in medical evidence. Clinical trials with unflattering results go unpublished, industry sponsorship corrupts the evidence base, studies suffer methodological flaws, and separate studies on the same intervention yield conflicting results. Says *Lancet* editor Richard Horton, "The case against science is straightforward: much of the scientific literature, perhaps half, may simply be untrue" (2015, 1380).

Recent years have seen the emergence of a new medical research domain known as 'meta-research' that studies these phenomena, among other research trends. As its figurehead John Ioannidis describes the field, "Meta-research is an evolving scientific discipline that aims to evaluate and improve research practices. It includes thematic areas of methods, reporting,

*To contact the author, please write to: Faculty of Medicine, University of Toronto, Toronto, Canada; African Centre for Epistemology and Philosophy of Science, University of Johannesburg, Johannesburg, South Africa; e-mail: jonathan.fuller@mail.utoronto.ca.

†Thanks to Mathew Mercuri, Nicolas Wuethrich, and audiences at the Philosophy of Science Association biennial meeting in Atlanta and the Institute for the History and Philosophy of Science and Technology in Toronto for productive feedback and discussion. I am grateful for funding support from the McLaughlin Centre.

Philosophy of Science, 85 (December 2018) pp. 767–780. 0031-8248/2018/8505-0003$10.00
Copyright 2018 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/699689 Published online by Cambridge University Press

reproducibility, evaluation, and incentives . . . helping science progress faster by conducting scientific research on research itself. This is the field of meta-research" (Ioannidis et al. 2015, 1–2). Ioannidis suggests that meta-research interfaces with many disciplines—including history and philosophy of science. In this article, I will explore one philosophical question about this hot new field: what to make of the relevance of its findings for clinical medicine.

Here are some notable examples of the kinds of results that the field is delivering. A meta-analysis (a pooled analysis of primary studies) found that industry-sponsored drug and device studies are 1.32 times as likely to report favorable efficacy results and 1.87 times as likely to report favorable harms results compared to non-industry-sponsored studies (Lundh et al. 2012), a result that is taken to be evidence of an 'industry bias' that distorts research findings. Another meta-analysis quantified the extent of publication bias, the preferential publication of studies with favorable results, and found that trials with positive results are 1.78 times as likely to be published compared to trials with negative results (Hopewell et al. 2009).

Many meta-research surveys measure the 'replicability' of studies or contradictions between studies. In a now-classic meta-survey, Ioannidis (2005) found that 41% of positive highly cited therapeutic studies that were compared with a second study of the same intervention were outright contradicted or had their effect estimate substantially revised downward by the second study. Pereira, Horwitz, and Ioannidis (2012) found that across the entire Cochrane Database of Systematic Reviews, when a very large treatment effect was measured in a first trial, 90% of the time a subsequent meta-analysis of trials substantially lowered the effect size estimate, and 34% of the time the meta-analysis found no statistically significant effect.

What is the value of these meta-research studies (beyond shocking and appalling their readership)? Ioannidis et al. claim, "While one can theorize about biases (e.g., publication bias, reporting bias, selection bias, confounding), it is now possible to examine them across multiple studies and to think about ways to prevent or correct them" in future research (2015, 2). But might it also be possible to use meta-research on these problems to correct for them in the therapeutic evidence we have already generated, to use meta-research as 'meta-evidence'?

Elsewhere, Ioannidis gestures toward this idea: to avoid being misled by inflated effect size estimates from early clinical research that are well documented in meta-research studies, he suggests we could consider "rational down-adjustment of effect sizes" or the use of "analytical methods that correct for anticipated inflation" (2008, 644). Pereira et al. (2012) hint that their meta-study could have implications for the reliability of evidence of very large treatment effects. Stegenga (2018) argues from the existence of widespread problems such as publication bias and discordant evidence to the nihilist conclusion

https://doi.org/10.1086/699689 Published online by Cambridge University Press

that on average we should have low confidence in the effectiveness of medical interventions, often citing meta-research in favor of his thesis. Could meta-research help in evaluating particular medical technologies?

The standard approach in evidence-based medicine (EBM) is to use only direct research evidence. For instance, the GRADE approach to evaluating therapeutic evidence assigns a level of confidence to an effect estimate from a therapeutic study or body of evidence by critically appraising study design and study results (Balshem et al. 2011). While GRADE does consider whether publication bias may have influenced the results, it assesses this possibility by attending to characteristics of the direct evidence, often applying graphical or statistical tests (Guyatt et al. 2011). My proposal extends the popular principle that medical judgments should be 'based on evidence' to the meta level.

In this article, I argue that meta-research findings should be used as meta-evidence for evaluating therapies. I start by introducing some distinctions among higher-order evidence, meta-evidence, and meta-research evidence. I then make the case for meta-research evidence, considering several important objections to its use in evaluating therapies: the irrelevance objection, the screening-off objection, and the underdetermination objection. I argue that meta-research evidence functions by rationally revising our confidence in first-order evidence and in the therapeutic hypothesis—typically, downward.

2. Higher-Order Evidence, Meta-Evidence, and Meta-Research Evidence. Evidence about evidence, or evidence about an agent's reasoning, is often called 'higher-order evidence' (HOE), and it presents puzzles to which epistemologists have recently turned their attention. It will be useful to see how well this concept of HOE captures the meta-research we are discussing.

Paradigm cases of HOE typically involve agents who reason from a body of evidence relevant to some hypothesis and then subsequently come across reasons to suspect that their initial reasoning was unreliable. Christensen (2010) discusses a hypothetical case he names 'Reasonable Prudence', in which a medical resident makes a diagnosis and prescribes a medication. The resident then realizes that he has been awake for 36 hours and knows that people tend to make errors when sleep-deprived. The resident perhaps even knows that he personally has a poor track record when he is so short on sleep. One question we can ask about a case like this one is, how should the resident regard his initial diagnostic and therapeutic conclusions after learning that his reasoning may have been impaired? Similar paradigm cases in the literature involve a 'Sleepy Detective' weighing evidence about a robbery during an all-nighter (Horowitz 2014), an anesthesiologist evaluating evidence about the optimal pain medication dose and then learning that he had been slipped 'reason-distorting mushrooms' ('Calculation'; Sliwa and Horowitz 2015), and an eye-

https://doi.org/10.1086/699689 Published online by Cambridge University Press

witness feeling very confident that she saw a particular suspect commit the murder before learning of the empirical evidence from psychology about the relative unreliability of eyewitness testimony (Roush 2009).

In these cases, the direct evidence about the hypothesis—diagnostic evidence, one's first-person recall of a crime—is the 'first-order evidence'. The evidence that the agent's first-order reasoning may be untrustworthy due to sleep deprivation, intoxication, or the frailty of human memory is HOE. Sliwa and Horowitz make the distinction in this way: first-order evidence bears directly on the hypothesis, while HOE "bears directly on the reliability of [the agent's] reasoning" (2015, 2836). Elsewhere, Horowitz describes HOE as "evidence about what evidence one has, or what one's evidence supports" (2014, 718).

In many of the representative cases and definitions, HOE concerns individuals, and it concerns either their reliability as epistemic agents or the correctness of their inferences from the first-order evidence. In Reasonable Prudence, Sleepy Detective, and Calculation, the HOE is relevant only to the sleep-deprived or intoxicated agent's reasoning; it would not apply to a cognitively unimpaired agent reasoning from the same first-order evidence. However, the meta-research with which we began is not evidence about a particular agent's reasoning; it concerns the public evidence from which many agents reason. Moreover, this meta-research is not straightforwardly evidence about the accuracy of inferences from the therapeutic studies in the same way that evidence about the track record of a doctor's diagnostic accuracy or the track record of eyewitness testimony measures the probability that a given diagnosis or testimony is accurate. Therefore, I will use the uncommitted term 'meta-evidence' to describe the kinds of paradigm meta-research findings we are discussing.

As a rough working concept, *first-order evidence* (FOE) E is direct evidence for the hypothesis H. In medicine, a clinical trial showing a positive result is FOE that the treatment is effective (H). Typically, when we speak of medical evidence we have some FOE in mind. FOE is 'direct' in comparison to meta-evidence. As a start, *meta-evidence* E' is evidence about E relevant to evaluating the evidential support that E lends to H, or how strongly E supports H.[1] A systematic review of publication bias is meta-evidence with respect to particular clinical trial evidence if it has some rational bearing on our evaluation of the support that the trial evidence lends to the hypothesis that the treatment is effective or safe. In what follows, I argue that meta-evidence is also (indirect) evidence with respect to the hypothesis.

1. I will not consider here how closely the paradigm examples of HOE fit into this rendering of meta-evidence. Some examples may fit better than others.

https://doi.org/10.1086/699689 Published online by Cambridge University Press

Finally, *meta-research evidence* is simply meta-evidence from meta-research. The distinction between meta-evidence and meta-research evidence is worth making because not all meta-evidence comes from meta-research. A physician could observe in her own experience that FOE in a certain domain often suffers from bias, which is plausibly meta-evidence about the FOE, though it is not meta-research evidence. Contrariwise, not all meta-research is meta-evidence because not all of it is relevant to assessing the evidential strength of FOE. My central claim is that meta-research evidence should be used in evaluating therapies, but in principle, other kinds of meta-evidence should be used as well.

3. The Case for Using Meta-Research as Meta-Evidence. Meta-research evidence is evidence about the support that first-order therapeutic studies (trials, systematic reviews, observational research) provide to the hypothesis that a treatment is effective or safe, or about their estimation of the therapeutic effect size. In this section, I argue that meta-research evidence on industry bias, publication bias, and contradictions in the medical literature (among other findings) should be used in evaluating therapies.

In section 3.1, I motivate my argument using three cases: Industry Bias, Publication Bias, and Contradiction. In sections 3.2–3.4, I develop the argument further by defending it against several objections: the irrelevance objection, the screening-off objection, and the underdetermination objection. In the process, I show that meta-research evidence works by rationally updating our confidence in the FOE.

3.1. Paradigm Cases. I will use three realistic cases to motivate my argument.

Industry Bias. A physician reads the report of a trial sponsored by a drug company that provides evidence (E_1) for H_1: the drug is efficacious compared to placebo. The physician gains high confidence in H_1. She then learns of the results of a meta-review on industry bias, finding that industry-sponsored studies like E_1 are 1.32 times as likely to be positive compared to non-industry-sponsored studies (Lundh et al. 2012). The physician becomes more confident that study E_1 is biased and less confident in H_1.

Publication Bias. A physician reads a meta-analysis of published trials (E_2) that finds a drug to be minimally beneficial (with a pooled result that is barely statistically significant), and he comes to believe H_2: that the drug is (minimally) efficacious. He then reads the meta-review by Hopewell et al. (2009) showing that while 73% of positive trials are published, only 41% of negative trials are published. The physician worries that there may be more negative unpublished trials compared to positive unpublished trials and that

https://doi.org/10.1086/699689 Published online by Cambridge University Press

selective publication may have biased the meta-analysis E_2 in favor of a positive pooled result. He becomes less confident in H_2.

Contradiction. A physician reads the results of a trial of a new drug (E_3) appearing to show that the drug has a very large effect size. She becomes confident in H_3: that the drug has a large positive effect. She then hears about two meta-research surveys conducted by Ioannidis's research group (Ioannidis 2005; Pereira et al. 2012) in which studies like this one are compared to superior follow-up studies. This meta-research shows that initial studies, especially those appearing to measure a very large effect, often fail to accurately predict the magnitude of effect in another study. The physician believes there is a good chance that E_3 fails to accurately estimate the effect size and becomes less confident in H_3.

In each case, the physicians start out with a high degree of confidence in the therapeutic hypothesis based on some FOE. They then come across meta-research relevant to evaluating the evidential support that the FOE lends to the hypothesis; namely, the meta-research raises the possibility that their FOE is systematically biased (I will discuss other interpretations of these meta-research findings in sec. 3.4 but will assume that they reveal bias in the meantime). In other words, the meta-research is meta-evidence. In response to the meta-research, the physicians rationally lower their confidence in the therapeutic hypothesis. On the principle that evidence is empirical information that leads us to rationally revise our confidence in the hypothesis, the meta-research was evidence with respect to the therapeutic hypothesis. Thus, on commonplace epistemic principles such as the principle of total evidence or the principle that one should respect one's evidence (Feldman 2005; Sliwa and Horowitz 2015), meta-research findings like these should be used in evaluating therapies.

I will now explore several illuminating objections to this argument.

3.2. Irrelevance Objection. The first objection, or pair of objections, argues that meta-research is irrelevant to the first-order therapeutic hypothesis. One variant of this objection worries that by changing our confidence in the hypothesis to suit the meta-evidence, we fail to respect the rational bearing of the FOE. Another variant notes that the hypothesis in question is causal; we must therefore settle the hypothesis through causal inference, and our theories of causal inference do not rely on anything like meta-research.

There is an apparent tension between meta-evidence and FOE. The FOE supports the first-order hypothesis with a certain level of confidence, while the meta-evidence is evidence that the FOE should support a different level of confidence.[2] In our paradigm cases, each physician became confident in

2. The paradoxical cognitive state of simultaneously believing both pieces of evidence is described in the HOE literature as 'epistemic akrasia' (Horowitz 2014).

https://doi.org/10.1086/699689 Published online by Cambridge University Press

the therapeutic hypothesis after learning about some first-order trial or meta-analysis but then came to believe that the trial or meta-analysis may have suffered from industry bias or publication bias (e.g.) and thus warrants a lesser confidence in the hypothesis. One way out of this tension is to drop one's confidence in the first-order hypothesis, in line with the higher-order information. The physicians in our cases all took this escape.

However, the irrelevance objection argues that in so doing they failed to respect the rational bearing of the FOE. We can safely assume that the results of a trial or meta-analysis are directly relevant to the hypothesis that the treatment works. Let us also assume that the physicians in all three cases correctly judged their FOE and were rational in having high confidence in the therapeutic hypothesis based on this evidence. So, the objection goes, for them to change their mind after learning about the meta-research was irrational because their new beliefs were no longer a good reflection of their FOE. Instead, they should have given up their meta-evidence.

One way to develop this objection further is to wonder, through what inference do therapeutic study results support the hypothesis? The irrelevance objection notes that the hypothesis in question concerns the effectiveness or effect size of a treatment, which are causal concepts. We must therefore decide the hypothesis through causal inference, and our best theories of causal inference do not rely on meta-research.

I think that this objection could call on any popular theory of causal inference, including those that use causal Bayes nets, counterfactual frameworks, or the probabilistic theory of causality. We can represent causal inference from trial results or other epidemiologic study results using the following highly abstracted form:

1. Metaphysical assumption(s).
2. Difference in outcome or probability of outcome between groups (from FOE).
3. The right kind of causal comparability between groups (from FOE).

H: The intervention caused the difference in outcome or probability of outcome.

Different theories formalize the premises differently. For instance, Cartwright's (2010) theory of the ideal randomized control trial uses the probabilistic theory of causality for the metaphysical assumption and understands causal comparability to mean an equal distribution of causally homogeneous subpopulations. My theory of the ideal comparative group study (Fuller 2018) formulates causal comparability as an equal contribution of a variable *C*, representing all the complex causal conditions.

However we choose to fill in the details, we need only FOE to support the second and third premises. In a therapeutic study, we measure the difference

https://doi.org/10.1086/699689 Published online by Cambridge University Press

in outcome between groups and assess causal comparability between groups by evaluating study design and analysis. The premises are engineered such that together they entail the conclusion. If we added another premise to this argument to represent the meta-research findings, it would not make the inference any more valid or sound. Nor is it easy to see how meta-research could provide confirmation or disconfirmation for the hypothesis independently of this argument because on its own meta-research does not provide the ingredients needed for a causal inference. So, the objection goes, meta-research is not relevant to the hypothesis.

My answer to the irrelevance objection is that the function of meta-research is not to provide independent confirmation or disconfirmation for H nor to strengthen or weaken the evidential relation between FOE and H, but to rationally revise our confidence in the FOE (and thus our confidence in the premises of our causal inference). In our paradigm cases, it lowered the physicians' confidence that the FOE was unbiased and thus lowered their confidence that the study groups were causally comparable in the right ways (for instance, perhaps bias introduced causally relevant baseline differences). By altering our confidence in the FOE, meta-evidence alters our confidence in the hypothesis that the FOE supports. If we are less certain that the study groups are causally comparable, we should also be less certain that the intervention caused the difference in outcome.

We can formalize my argument using probabilities to represent our confidence. We can let $p(H)$ be our confidence in H, $p(H|E)$ be our confidence in H given positive (unbiased) FOE, $p(H|\neg E)$ be our confidence in H in the absence of the FOE, $p(E)$ be our confidence that the (unbiased) evidence obtains, and $p(\neg E)$ be our confidence that the evidence does not obtain. On the total probability equation, $p(H) = p(H|E)p(E) + p(H|\neg E)p(\neg E)$. Because E confirms H, $p(H|E) > p(H|\neg E)$. Thus, whenever meta-research evidence E' changes our confidence in E (the $p(E)$), we must update our confidence in H (the $p(H)$) to avoid being irrational.

This account of how meta-evidence works addresses the concern raised earlier (and in the HOE literature by Sliwa and Horowitz [2015]) that by altering our confidence in H, we fail to respect E's bearing on H. The 'rational bearing of E on H' is represented by $p(H|E)$ and can be stronger or weaker depending on the first-order inference we employ (for our causal inference, it is the probability of the conclusion given positive unbiased study results). Meta-evidence acts through $p(E)$ rather than $p(H|E)$. Meta-evidence may be irrelevant to the bearing of E on H, but it is entirely relevant to our confidence in E and thus to our confidence in H.

We can now be more precise in describing how meta-research—as meta-evidence—is 'relevant to the evidential support that FOE lends to H'. It is relevant to evaluating our confidence in FOE, and our confidence in FOE is relevant to evaluating the degree to which FOE confirms H. In contrast,

https://doi.org/10.1086/699689 Published online by Cambridge University Press

accounts of HOE have understood HOE as an 'undercutting defeater' that severs the logical connection between FOE and conclusion (Feldman 2005) or as compelling us to bracket the justification provided by FOE (Christensen 2010).

In summary, the irrelevance objection worries that meta-research is irrelevant for evaluating therapies because the rational bearing of FOE trumps meta-evidence; and meta-research findings do not play a role in the causal inferences that FOE supports. But meta-research evidence functions by rationally revising our confidence in the FOE and thus in the premises of our causal inference rather than by disrespecting the rational bearing of FOE on the hypothesis.

3.3. Screening-Off Objection. The next objection worries that meta-research investigates a class of FOE rather than the FOE token in question. Instead of relying on meta-research findings, we should carefully assess the FOE token for signs of bias. This assessment screens off the meta-research evidence.

To elaborate, meta-research on industry bias, publication bias, or contradictions in the medical literature might reveal the rate of bias or inaccuracy in the evidence base, which could help us to estimate the probability that a particular piece of FOE is biased, given that it comes from our evidence base. But the evidence base is a broad and heterogeneous reference class. Instead of relying on the probability of bias given this broad reference class, we should assess the probability of bias in the ideal reference class formed by the FOE token, the particular trial or meta-analysis. The information about this much narrower reference class 'screens off'—renders needless—the information provided by meta-research about the broader class. The physicians in our cases should have determined how confident they could be in the FOE by evaluating it using traditional EBM critical appraisal, as GRADE recommends. To assess the likelihood of publication bias, they could use trial registries, funnel plots, or statistical tests. To assess the likelihood of industry bias, they could examine the study design and results for signs of manipulation.

The problem with this line of objection is that it is a common feature of the kinds of problems that meta-research analyzes that are difficult to detect in the individual case. The bias is often hidden or not obvious. The mechanisms of industry bias are varied (Sismondo 2008) and are not always apparent to the critical appraiser. They can involve subtle choices that tip the results in a favorable direction and that are not reported or that have a difficult-to-determine influence. In fact, in their systematic review, Lundh et al. (2012) conclude, "Our analyses suggest the existence of an industry bias that cannot be explained by standard 'risk of bias' assessments" (2) and argue from their data that "an assessment of [industry] sponsorship should therefore be used as a proxy for these mechanisms" of industry bias (15).

https://doi.org/10.1086/699689 Published online by Cambridge University Press

For publication bias, trial registries are incomplete (Hopewell et al. 2009) and often fail to alert us to unpublished studies. While graphical and statistical tests of publication bias may be useful, the GRADE group concedes that they are error-prone and rely on strong assumptions and that publication bias is "difficult to predict for individual systematic reviews" (Guyatt et al. 2011, 1278).

Thankfully, we have meta-evidence that measures the frequency and severity of these problems in the evidence base. Of course, when we can better judge their influence on an individual evidence token, we should base our confidence in the FOE on this judgment. The screening-off objection holds that our appraisal of the FOE always screens off the meta-research findings. But screening off occurs only when we can satisfactorily judge our confidence in the FOE just using traditional critical appraisal. Oftentimes, we also need meta-research evidence.

3.4. Underdetermination Objection. The final objection I will discuss could grant that on the interpretations of meta-research findings with which we have been working (i.e., that they are evidence of bias in FOE), they warrant a lower confidence in the hypothesis; but in practice meta-research is tough to interpret. There are many possible readings, not all of which would support a lowering of our confidence, and none of which straightforwardly provides us with a revised quantitative level of confidence. Meta-research underdetermines our confidence in therapies and thus should not be used in evaluating them.

In Industry Bias and Publication Bias, the physicians inferred that the meta-research findings were evidence of bias and downgraded their confidence in the drug accordingly. But, one might worry, it could be that some other explanation truly accounts for the association between private funding and positive study results. In the Contradiction case, perhaps the physician was similarly hasty because there are many possible reasons beyond bias for why different studies of the same intervention might fail to agree. The correct explanation is underdetermined by the meta-research findings themselves, which simply measure the association between industry funding and study results, the association between publication and study results, or the rates of contradiction between studies. So perhaps the rational response for these physicians would be to suspend judgment with respect to the meta-research and stick to their initial evaluation based solely on the FOE.

I could retreat at this point and argue that, in principle, if we could attribute these meta-research trends to bias, then we should use them in evaluating therapies as the physicians in our simplified cases. Instead, I will argue that although there are several plausible interpretations of these meta-research findings, all of them should lead us to either lower or maintain our initial confidence in the FOE. Thus, after we allocate some credence to each of the plau-

https://doi.org/10.1086/699689 Published online by Cambridge University Press

sible interpretations, on net they should lead us to lower our confidence in the FOE.

Take the meta-research on 'publication bias'. It could be that negative studies are just as high in quality as positive studies, in which case preferentially failing to publish the negative ones asymmetrically removes some of the true variability and statistical noise in the data, biasing our pooled estimate toward a more positive value. As a second possibility, maybe negative studies are more likely to have flaws or small sample sizes (perhaps that is why they are not published), and thus excluding them from our published evidence base results in a more accurate pooled estimate. Because there is no research supporting an association between study quality and published results and some evidence that published and unpublished studies have similar sample sizes (Hopewell et al. 2009), we should regard the first possibility as a far more plausible interpretation. (A third possibility, that positive studies tend to have a lower quality, would make selective publication of positive studies even more biasing.) Therefore, the meta-evidence of publication bias should diminish our confidence in a positive systematic review.

There are two plausible explanations for the association between industry sponsorship and study results. It could be that drug and device companies tend more often to study effective and safe treatments compared to non-industry-sponsored researchers—there is no bias here. Or (more plausibly), industry-sponsored studies tend, on average, to overestimate effectiveness through various mechanisms (Lundh et al. 2012)—'industry bias'. A third explanation, 'nonindustry bias', holds that non-industry-sponsored studies tend to underestimate effectiveness; but this explanation is far less plausible than industry bias because mechanisms of industry bias are widely documented (Sismondo 2008), while mechanisms of nonindustry bias are not. So allocating some credence to the two plausible explanations, on net we should lower our confidence in a positive industry-funded study in response to the meta-evidence.

When considering meta-research on rates of contradiction in the medical literature, the number of plausible interpretations multiplies. 'Nonreplication explanations' chalk the disagreement up to a failure to replicate the results of the same intervention, given the same relevant causal factors, due to either a spurious chance finding or bias in one or both studies. Meanwhile, 'nontransportability explanations' locate the disagreement in the nontransportability of the true effect between two populations that differ in relevant causal factors. If the hypothesis is that your therapeutic study shows efficacy for the study population, the meta-evidence on contradiction rates should lower your sum confidence because nonreplicability explanations are confidence lowering, while nontransportability explanations are confidence neutral. If the hypothesis is instead that your therapeutic study predicts effectiveness for some target population outside the study, the meta-evidence should similarly lower

https://doi.org/10.1086/699689 Published online by Cambridge University Press

your confidence because both types of explanations are confidence lowering: neither falsely positive results nor truly positive, nontransportable results predict effectiveness elsewhere.

Another variant of the underdetermination objection argues that even if I am right that meta-research evidence should lower our credence in *E* and *H*, it is difficult to quantify how much lower our credence should be; so we should not use meta-research as meta-evidence because its precise implications are difficult to discern. However, current approaches to evaluating our confidence in therapies are qualitative to begin with (Balshem et al. 2011). In the very least, meta-research should qualitatively lower our confidence in FOE and the therapeutic hypothesis. Sometimes, knowing that our confidence in therapeutic effectiveness should be lower is informative, especially when the prior expectation of benefit barely outweighs the expectation of harm and a shift in our confidence might tip the balance of expectations in the opposite direction. Moreover, though it may be difficult, we can and should quantitatively model the evidential import of meta-evidence, as some philosophers have done with meta-research findings or HOE (Roush 2009; Sliwa and Horowitz 2015; Stegenga 2018).[3]

The underdetermination objection may be right that our paradigm meta-research findings admit to multiple interpretations and have imprecise evidential import, but on net the plausible interpretations compel us to lower our confidence in therapies, at least qualitatively. Therefore, this objection is not an impregnable barrier to using meta-research evidence.

4. Conclusion. Meta-research findings should be used as meta-evidence for evaluating therapies. Meta-evidence is evidence about the support that direct FOE lends to a hypothesis. Meta-research evidence fills this role by rationally revising our confidence in FOE and, consequently, in the hypothesis. I considered several objections to using meta-research evidence for evaluating therapies. First, the irrelevance objection argues that meta-research findings are not relevant to the therapeutic hypothesis; only FOE has rational bearing on this causal matter. But meta-research evidence lowers our credence in the premises of our causal inference, while preserving the justification that the premises provide the hypothesis. Next, the screening-off objection says that critically appraising the FOE screens off the meta-evidence, rendering meta-evidence needless. Unfortunately, the problems that meta-research ana-

3. Bayesian apparatus may be particularly helpful here given the natural tendency to regard meta-evidence as informing second-order degrees of belief. Stegenga (2018) uses Bayes's rule to model the influence of meta-research on our confidence in therapeutic effectiveness, while Landes, Osimani, and Poellinger (2017) develop the Bayesian modeling approach of Bovens and Hartmann (2003) for the assessment of therapeutic harms, allowing them to account for a therapeutic study's reliability.

https://doi.org/10.1086/699689 Published online by Cambridge University Press

lyzes are often difficult to detect in a FOE token, so we must often rely on meta-evidence for a class of FOE. Finally, the underdetermination objection argues that meta-research findings underdetermine our confidence in FOE and in the hypothesis because there are many possible interpretations of meta-research findings and they offer no definite quantitative prescriptions. Yet, on average, the plausible interpretations should lower our confidence in FOE and in the hypothesis, at least qualitatively. How best to model the evidential import of meta-research evidence in evaluating therapies remains an open question.

Some physicians and evidence users may already lower their confidence in therapies based on the ominous emerging meta-research evidence. However, my proposal departs from the explicit recommendations for critical appraisal in evidence-based medicine. Nonetheless, I see my proposal as an extension of a widely accepted EBM principle—that medical judgment should be based on evidence—to the next level: the meta level.

REFERENCES

Balshem, Howard, et al. 2011. "GRADE Guidelines: 3. Rating the Quality of Evidence." *Journal of Clinical Epidemiology* 64 (4): 401–6.

Bovens, Luc, and Stephan Hartmann. 2003. *Bayesian Epistemology*. Oxford: Oxford University Press.

Cartwright, Nancy. 2010. "What Are Randomised Controlled Trials Good For?" *Philosophical Studies* 147 (1): 59–70.

Christensen, David. 2010. "Higher-Order Evidence." *Philosophy and Phenomenological Research* 81 (1): 185–215.

Feldman, Richard. 2005. "Respecting the Evidence." *Philosophical Perspectives* 19:95–119.

Fuller, Jonathan. 2018. "The Confounding Question of Confounding Causes in Randomized Trials." *British Journal for the Philosophy of Science*, online first. doi:10/1093/bjps/axx015.

Guyatt, Gordon H., et al. 2011. "GRADE Guidelines: 5. Rating the Quality of Evidence—Publication Bias." *Journal of Clinical Epidemiology* 64 (12): 1277–82.

Hopewell, Sally, Kirsty Loudon, Mike J. Clarke, Andrew D. Oxman, and Kay Dickersin. 2009. "Publication Bias in Clinical Trials Due to Statistical Significance or Direction of Trial Results." *Cochrane Database Systematic Reviews* 1:MR000006.

Horowitz, Sophie. 2014. "Epistemic Akrasia." *Nous* 48 (4): 718–44.

Horton, Richard. 2015. "Offline: What Is Medicine's 5 Sigma?" *Lancet* 385 (9976): 1380.

Ioannidis, John P. 2005. "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research." *JAMA* 294 (2): 218–28.

———. 2008. "Why Most Discovered True Associations Are Inflated." *Epidemiology* 19 (5): 640–48.

Ioannidis, John P., Daniele Fanelli, Debbie Drake Dunne, and Steve N. Goodman. 2015. "Meta-Research: Evaluation and Improvement of Research Methods and Practices." *PLoS Biology* 13 (10): e1002264.

Landes, Jurgen, Barbara Osimani, and Roland Poellinger. 2017. "Epistemology of Causal Inference in Pharmacology: Towards a Framework for the Assessment of Harms." *European Journal for Philosophy of Science* 8:3–49.

Lundh, Andreas, Joel Lexchin, Barbara Mintzes, Jeppe B. Schroll, and Lisa Bero. 2012. "Industry Sponsorship and Research Outcome." *Cochrane Database Systematic Reviews* 12:MR000033.

Pereira, Tiago V., Ralph I. Horwitz, and John P. Ioannidis. 2012. "Empirical Evaluation of Very Large Treatment Effects of Medical Interventions." *JAMA* 308 (16): 1676–84.

https://doi.org/10.1086/699689 Published online by Cambridge University Press

Roush, Sherrilyn. 2009. "Second Guessing: A Self-Help Manual." *Episteme* 6 (3): 251–68.
Sismondo, Sergio. 2008. "How Pharmaceutical Industry Funding Affects Trial Outcomes: Causal Structures and Responses." *Social Science and Medicine* 66 (9): 1909–14.
Sliwa, Paulina, and Sophie Horowitz. 2015. "Respecting *All* the Evidence." *Philosophical Studies* 172 (11): 2835–58.
Stegenga, Jacob. Forthcoming. *Medical Nihilism*. Oxford: Oxford University Press.

https://doi.org/10.1086/699689 Published online by Cambridge University Press