



Kent Academic Repository

Wilde, Michael Edward (2023) *The EBM+ Movement*. International Journal of Biostatistics, 19 (1). pp. 283-293. ISSN 1557-4679.

Downloaded from

<https://kar.kent.ac.uk/99089/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1515/ijb-2022-0126>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

The EBM+ Movement

Michael Wilde. Forthcoming in *The International Journal of Biostatistics*.

Abstract

In this paper, I provide an introduction for biostatisticians and others to some recent work in the philosophy of medicine. Firstly, I give an overview of some philosophical arguments that are thought to create problems for a prominent approach towards establishing causal claims in medicine, namely, the Evidence-Based Medicine (EBM) approach. Secondly, I provide an overview of further recent work in the philosophy of medicine, which argues that mechanistic studies can help to address these problems. Lastly, I describe a novel approach for establishing causal claims in medicine that has been informed by this recent work in the philosophy of medicine, namely, the EBM+ approach.

Introduction

*How do we establish causal claims about the effectiveness of medical interventions? One approach is to carry out some Evidence-Based Medicine (EBM). Of course, most people agree that such causal claims are established on the basis of the evidence; the devil is in the details (cf. Worrall 2007: 981-2). In this paper, I begin by providing some of the details of the EBM approach. In particular, I will present its emphasis on the role of clinical studies in establishing causal claims about the effectiveness of medical interventions. Next, I present a couple of problems with this approach, emphasized by recent work in the philosophy of science, for instance, due to Nancy Cartwright (2007, 2010) and John Worrall (2002, 2007). I then provide an overview of further recent work in the philosophy of science to the effect that mechanistic studies can help to address these problems (Clarke *et al.* 2014). And I will*

describe a novel approach for establishing causal claims in medicine that has been informed by this recent work in the philosophy of science: the EBM+ approach (Parkkinen *et al.* 2018; Williamson 2019, 2021). I hope this paper serves as an introduction for biostatisticians and others to some of the implications of recent work in the philosophy of science, and in particular the philosophy of medicine.

EBM

According to a classic definition, EBM is ‘the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients’ (Sackett *et al.* 1996: 71, quoted by Clarke *et al.* 2014: 339). I will focus on decisions about the best treatment of patients, where a good treatment decision often involves establishing a causal claim about the effectiveness of a medical intervention, and then integrating this knowledge with the individual patient’s values and circumstances (Howick 2011a: 177). Presumably, any clinician worth their salt has always believed themselves to be making such decisions on the basis of a conscientious, explicit, and judicious use of the current best evidence concerning the effectiveness of medical interventions. EBM is novel in: (i) often appealing to the notion of an evidence hierarchy in order to explain what counts as the current best evidence; (ii) providing explicit guidance to help ensure that this current best evidence is used conscientiously, explicitly, and judiciously.

An evidence hierarchy is simply a ranking of different methods for providing evidence to inform medical decisions: loosely speaking, the higher the ranking of the method, the better the evidence that it provides. In the case of treatment decisions, an evidence hierarchy will typically rank methods in accordance with their potential for establishing a causal claim

about the effectiveness of a medical intervention, at least all other things being equal. Such a hierarchy ranks comparative clinical studies above mechanistic reasoning, reflecting the view that clinical studies are better than mechanistic reasoning at establishing causal claims about the effectiveness of medical interventions (Howick 2011a: 5). This evidence hierarchy helps to explain what counts as the current best evidence: if there are clinical studies available, then they provide the current best evidence, otherwise ‘we must follow the trail to the next best external evidence and work from there’ (Sackett *et al.* 1996: 72). And if we follow the trail all the way down, then we end up with expert judgement and mechanistic reasoning.

Although in practice this next best external evidence is rarely taken to include non-clinical studies; indeed, even non-randomized clinical studies are sometimes not included in the next best external evidence: ‘[i]f the study wasn’t randomized, we’d suggest that you stop reading it and go on to the next article in your search’ (Sackett *et al.* 2000: 108).

What exactly is mechanistic reasoning? It perhaps helps to begin with the standard definition of a mechanism from the philosophy of science: a *mechanism* is something that can be appealed to in an attempt to explain some phenomenon; it is simply an arrangement of entities that perform certain activities in order to produce that phenomenon (Machamer *et al.* 2000: 3). For example, if we want to explain the putative effectiveness of some medical intervention, then we can point towards its proposed mechanism of action, such as the proposed mechanism by which inserting a small valve or grommet in the eardrum works to equalize pressure in the ear, allowing the build-up of glue-like fluid in the middle of the ear canal to be drained away, thereby improving so-called *glue ear* (Worrall 2007: 985).

Mechanistic reasoning involves inferring a causal claim about the effectiveness of a medical intervention simply on the basis of a claim about the existence of such a mechanism (Auker-Howlett and Wilde 2020; Howick 2011a; Wilde 2021). To continue the present example, on

the basis of the claimed mechanism linking the insertion of grommets to the draining of fluid in the middle ear, one may infer the causal claim that grommets are an effective intervention for the treatment of patients with glue ear.

One worry is that mechanistic reasoning is known to have a pretty bad track record. Indeed, Howick (2011a: 154-7) lists a number of cases where ineffective or harmful medical interventions were recommended on the basis of mechanistic reasoning. For example, it is now thought that the insertion of grommets offers no benefit for the treatment of glue ear (Worrall 2007: 985). Moreover, there is a good explanation of this bad track record for mechanistic reasoning. Firstly, such reasoning may suffer from *the problem of story-telling*: mechanistic reasoning can be overly psychologically compelling, discouraging further investigation simply because people are too easily convinced by a gripping narrative (cf. Clarke *et al.* 2014: 350). Mechanistic reasoning is therefore liable to be based upon a false theory about the details of some mechanism, in which case such reasoning is not going to establish a causal claim about the effectiveness of the relevant medical intervention.

Secondly, mechanistic reasoning suffers from *the problem of incompleteness*: the typical complexity of the relevant mechanisms means that any knowledge of the details of those mechanisms is often incomplete; even if we have some knowledge of an intervention's mechanism of action, the intervention may interact with known mechanisms in unexpected ways, or it may set off other unknown mechanisms that work to counteract or mask the expected effects of the intervention, such that the intervention has no overall beneficial effect (Howick 2011b; Illari 2011; Wilde 2021). It is therefore difficult to obtain the knowledge required for mechanistic reasoning to provide enough evidence to establish a causal claim about the effectiveness of a medical intervention. This problem with mechanistic reasoning have been nicely summed up by the philosopher of science Miriam Solomon: 'A general

problem with mechanistic accounts is that they are typically incomplete, although they often give an *illusion* of a complete, often linear, narrative' (2015: 131).

A proponent of the above evidence hierarchy will prefer that treatment decisions are based upon on the results of comparative clinical studies, such as observational studies or preferably randomized controlled trials (cf. Guyatt *et al.* 1992: 2420, quoted by Williamson 2021: 192).

A standard type of clinical study involves comparing pre-determined average health outcomes of participants with a certain disease sometime after they have been separated into one of two groups, where those participants in the test group received some novel intervention, and those participants in the control group received the current standard treatment; if there is a difference in such a health outcome observed between the test group and the control group, then the existence of a correlation is established between the intervention and that health outcome; and if the novel intervention is the only difference between the test group and the control group, then it must be the intervention that explains this correlation, that is, a causal claim about the effectiveness of the intervention is established. For example, a clinical study may reveal that a novel intervention is correlated with an improved average time taken to recover from some disease; if that intervention is the only difference between the test group and the control group, then the study establishes not only the claim that the intervention is correlated with improved recovery times, but also the causal claim that the novel intervention is effective at reducing the average time taken to recover from the disease (cf. Cartwright 2007).

Of course, the difficulty is in ensuring that the test group and the control group differ only with respect to the novel intervention; if there remains some other difference between the groups, then it may be that any correlation between the novel intervention and some health

outcome is explained by appealing to this other difference, at least as long as this other difference is a factor plausibly relevant to the prognosis of the disease, that is, as long as the difference concerns a so-called *prognostic factor*. For example, if the participants in the test group were aware that they were receiving a novel intervention, or if these participants were treated differently to those participants in the control group, then it may be these differences that explain any established correlation. These sorts of differences can be addressed by carrying out double-blinded, placebo-controlled trials, in strict accordance with a sensible study protocol (cf. Cartwright 2011a: 1400). However, if the control group still differs from the test group in terms of some prognostic factor, then it may be this difference that explains any established correlation between the intervention and some health outcome in the study. For example, the control group may perform less exercise on average than the test group, and it may be this difference that explains any established correlation between the intervention and improved average time taken to recover from the disease; that is, level of exercise is said to be a potential *confounding factor*, that is, a factor other than the intervention that may be appealed to in order to explain an established correlation between the intervention and some health outcome. As the old saying goes: correlation is not causation.

One method to address this *problem of confounding factors* involves intentionally matching the test group and the control group in terms of such factors, for example, by ensuring that both groups perform on average the same level of exercise. However, it is only possible to carry out this matching with respect to known potential confounding factors; after such matching, there may remain unknown confounding factors, that is, unsuspected prognostic factors that may be appealed to in order to explain an established correlation between the intervention and some health outcome (Worrall 2002: S322; Worrall 2007: 1003-4). So the preferred method involves allocating participants in the study to the test or control group at

random, that is, by carrying out a randomized controlled trial. The idea is that the chance that a participant with a certain prognostic factor will be included in the test group is equal to the chance that they will be included in the control group. As a result, if the trial is sufficiently large, then there is a high chance that this factor will be evenly distributed between the test group and the control group. For example, following randomization, a sufficiently large trial will result in a high chance that the test group and the control group perform on average the same level of exercise, even if those groups have not been deliberately matched for this prognostic factor. And so any significant difference in average health outcomes between the test group and the control group cannot be explained by appealing to a difference in level of exercise. Indeed, following randomization, a sufficiently large trial will result in a high chance that any given prognostic factor is similarly distributed in the test group and the control group. And so if such a randomized trial establishes the existence of a correlation between a novel intervention and some health outcome, the trial also arguably establishes the causal claim that it was the intervention that caused this health outcome, on the grounds that the correlation cannot be explained by appealing to some confounding factor. It is in this sense that a randomized trial is said to control for all potential confounding factors, known or unknown. And it is partly for this reason that randomized trials are ranked above non-randomized observational studies in the evidence hierarchy for treatment decisions (Howick 2011a: 31-116). Of course, things are not quite so cut and dried: more recent evidence hierarchies allow for the possibility that a sufficiently high quality observational study can provide better evidence than a sufficiently low quality randomized trial (Guyatt et al. 2008). And evidence hierarchies more generally permit a number of different interpretations (Jerkert 2021). Although, all other things being equal, randomized trials still rank higher than observational studies in these evidence hierarchies.

One of the putative advantages of a randomized trial is thus that it can establish a causal claim about the effectiveness of a novel intervention without requiring much in the way of knowledge about the exact mechanisms by which the intervention is effective. Accordingly, Richard Ashcroft says that:

[T]he beauty of the [randomized controlled trial] as a methodology is that it seems to operate at a level of scientific theory autonomous from the basic sciences. Apparently we need to know little or nothing of pathogenesis or drug action in order for a randomized controlled trial to be designed and implemented and (perhaps) interpreted successfully. Indeed, our theories at this more basic level could simply be wrong [or at least incomplete]. (2004: 134, quoted by Solomon 2015: 110)

It is for this reason that randomized trials are also ranked above mechanistic reasoning in the evidence hierarchy: randomized trials can establish the effectiveness of an intervention, while avoiding the problems associated with mechanistic reasoning, namely, the problem of incompleteness and the problem of story-telling. At least all other things being equal, a randomized trial thus provides the current best evidence about the effectiveness of a medical intervention.

Some philosophers of science have been critical of this emphasis on randomized trials; they have argued that randomized trials have their own set of problems. One such problem is sometimes known as *the problem of non-causal correlations* (Clarke *et al.* 2014: 343-44). A version of this problem has been explained in detail by John Worrall (2002, 2007), building upon work by Peter Urbach (1985, 1993). Worrall argues that randomization ensures at best a high chance that a given confounding factor is distributed similarly in the test and the control

group; there will remain a non-zero chance that some factor is better represented in one group than the other (2002: 322-3). He also argues that such a high chance is ensured only by indefinite randomization:

[I]f the randomization were performed indefinitely often, the number of cases of groups skewed with respect to that factor would be very small. The fact, however, is that a given [randomized trial] has not been performed indefinitely often but only once. Hence it is of course possible that, “unluckily,” the distribution even of the one unknown prognostic factor we are considering is significantly skewed between the two groups. (2002: S323)

In other words, if a correlation between an intervention and some health outcome is established in a randomized trial, it may not be because the intervention caused this health outcome, but rather because there was some other difference between the test and control groups that was not controlled for by randomization.

Now, it might be objected that a systematic review or meta-analysis of multiple replications of a randomized trial will provide an adequate approximation of indefinite randomization. In that case, there will be a high chance that any given prognostic factor is similarly distributed in the test and control groups. Indeed, in more detailed evidence hierarchies, systematic reviews and meta-analyses of randomized trials rank even higher than randomized trials alone (cf. Clarke *et al.* 2014: 340). However, Worrall then argues that it would be a quantification fallacy to conclude that any difference observed between the test and control group must then be attributable to the intervention (2002: S324). An example from contemporary epistemology may help to make clear this fallacy, namely, the example of a large lottery with

one winning ticket: although for each given ticket, there is a high chance that the ticket will lose, it does not follow that there is a high chance that every ticket will lose; indeed, the description of the example means that there is a high chance that some ticket will win (cf. Wheeler 2007). Similarly, although for each potential confounding factor, there is a high chance that it is distributed similarly in the test and control groups, it does not follow that there is a high chance that every such factor is distributed similarly in the new treatment and control treatment groups; there may even be a high chance that some confounding factor is not distributed similarly in the two groups, in which case even a correlation observed across multiple randomized trials may not be sufficient for establishing a causal claim about the medical intervention, since the correlation may be explained by appealing to the chance of there being such a factor. One possible example here concerns a meta-analysis of randomized trials on the clinical efficacy of homeopathy; this meta-analysis established that homeopathic interventions are correlated with better health outcomes than placebo, but this alone was not sufficient to establish that the interventions caused those better health outcomes in the study population, presumably because there was still a significant chance that there was a non-causal explanation of this correlation (Cucherat *et al.* 2000, cited by Clarke *et al.* 2014: 344). It is thus argued that the problem of non-causal correlations means that even a systematic review or meta-analysis of randomized trials may fail to establish an efficacy claim, that is, the claim that a medical intervention caused some health outcome in the study population. More generally, Jacob Stegenga (2011) has argued that meta-analyses should not be regarded as the platinum standard of evidence. In particular, Stegenga argues that a given meta-analysis will be insufficiently objective, since it involves making a number of subjective decisions, for example, decisions about which trials to include, and what weighting to give those trials in the final analysis.

It may be that the mere theoretical chance of there being a confounding factor is not sufficient by itself to preclude a systematic review or meta-analysis of randomized trials from establishing an efficacy claim about a medical intervention in practice (cf. Solomon 2015: 134-7). However, there remains another problem with emphasizing randomized trials, a problem that has been highlighted by Nancy Cartwright (2007, 2010). It is sometimes called *the problem of external validity* (Clarke *et al.* 2014: 346). A study has external validity to the extent that its results can be extrapolated from the study population to some target population. Cartwright allows that ideal randomized trials may indeed establish *the efficacy claim* that an intervention caused some health outcome in the study population, but then argues that this alone does not establish the causal claim of most interest, namely, *the effectiveness claim* that the intervention causes this health outcome for distinct target populations; indeed, the study population may not be representative of a target population, for instance, if participants were recruited to the study from the target population in strict accordance with narrow exclusion criteria. Indeed, trials may exclude participants with multiple comorbidities from the study population, even though the target population is likely to have such comorbidities (Clarke and Russo 2017). One example here concerns the randomized trials which established in their study populations that benoxaprofen was efficacious for musculo-skeletal pain; benoxaprofen was found not to be an effective treatment of musculo-skeletal pain in an older target population (Worrall 2007: 994-5). Cartwright argues more generally that ‘[t]he lesson to be learned is that although (ideal) [randomized trials] are excellent at securing causal principles, there is a very great deal more that must be assumed—and defended—if the causal principles are to be exported from the experimental [study] population to some target population’ (2010: 66).

Now, it might be argued that the problem of external validity does not always preclude randomized trials from establishing an effectiveness claim. Indeed, Mark Petticrew and Iain Chalmers (2011) have objected that the default presumption should be that a systematic review of good randomized trials on varied enough study populations does indeed have external validity. However, Cartwright has responded that this objection involves giving up evidence-based medicine in place of default-based medicine. She argues that '[w]e should not adopt default positions; we should have evidence. Very often we cannot get it and have to bet. But at that point, when evidence stops and betting starts, we are no longer doing evidence-based medicine' (2011b: 1697).

To sum up, a number of philosophers of science have argued that even a group of well-conducted randomized trials can be insufficient to establish a causal claim about the effectiveness of a medical intervention on some health outcome. Firstly, the trials may demonstrate only that the intervention is correlated with the health outcome, thanks to randomization failing to completely control for confounding factors; this is the problem of non-causal correlations. Secondly, even if the trials can in fact demonstrate that the intervention causes the health outcome in the study population, that is not to say that the intervention will similarly cause the health outcome in a distinct target population; this is the problem of external validity. *How then do we establish causal claims about the effectiveness of medical interventions, when clinical studies alone are insufficient?*

EBM+

Some philosophers of science have argued that establishing a causal claim about a medical intervention involves also establishing the existence of a relevant mechanism (Clarke *et al.*

2014; Parkkinen *et al.* 2018; Russo and Williamson 2007; Williamson 2019). One line of argument draws upon the nine viewpoints for distinguishing mere correlation from causation advocated by Austin Bradford Hill (1965). Among other things, Hill argues that, in determining whether an observed association is causal, it can be helpful to consider not just the *strength* of the observed association, but also its *coherence* with ‘the generally known facts of the natural history and biology of the disease’ (1965: 298). Federica Russo and Jon Williamson (2007: 160-1) have argued that the former *strength* viewpoint demonstrates the importance of establishing a correlation when establishing a causal claim, whereas the latter *coherence* viewpoint demonstrates the importance of also establishing the existence of a relevant mechanism. Another line of argument appeals to the uses of causal claims in medicine (Clarke *et al.* 2014: 345-6; Russo and Williamson 2007: 159). Causal claims are used in medicine for prediction and explanation; for instance, after establishing the claim that a medical intervention causes some health outcome, one is better placed to predict and explain that health outcome in the presence of that intervention. Russo and Williamson (2007) argue that such a causal claim can be useful for prediction only if the intervention is appropriately correlated with the health outcome, that is, only if the causal claim tracks the existence of an appropriate correlation. They argue also that such a causal claim can be useful for explanation only if the intervention is linked by a relevant mechanism to the health outcome, that is, only if the causal claim tracks the existence of a relevant mechanism. Their conclusions are sometimes summed up as *the Russo-Williamson thesis* (Gillies 2019: 133-84). A recent statement of the thesis goes like this: ‘In order to establish a causal claim in medicine one normally needs to establish two things: first, that the putative cause and effect are appropriately correlated; second, that there is some mechanism which explains instances of the putative effect in terms of the putative cause and which can account for this correlation’ (Williamson 2019: 33). Given this, it is perhaps unsurprising that establishing the

existence of a mechanism can help to address the problem of non-causal correlations and the problem of external validity.

Let us look first at the problem of non-causal correlations. Suppose that randomized trials have established a correlation between some novel intervention and an improved health outcome, but that the effectiveness of this intervention is not the only available explanation of this established correlation. In particular, it may be that the test and control groups still differ in some way other than the novel intervention, and this difference provides an alternative potential explanation of the extent of the established correlation. Attempting to establish a mechanism can help in deciding between these competing explanations (Clarke *et al.* 2014: 343-4). On the one hand, establishing the existence of a mechanism linking the novel intervention and the improved health outcome can help to rule out the alternative potential explanations, at least where this mechanism explains the extent of the established correlation. A potential example here involves the move to a *pegylated* combination therapy for chronic Hepatitis C (Auker-Howlett and Wilde 2020). Pegylation is the process of modifying molecules with polyethylene glycol (PEG) (Strader *et al.* 2004). Although randomized trials established a correlation between a pegylated combination therapy and improved health outcomes compared to the standard combination therapy, there remained a worry that this correlation could be explained by bias in the available trials rather than the improved effectiveness of the pegylated combination therapy. But independent evidence helped to rule out this alternative potential explanation, by helping to establish the existence of the antiviral mechanism of pegylation, where this mechanism could explain the extent of the established correlation. On the other hand, if the existence of such a mechanism was not established after such an investigation, then the effectiveness of the medical intervention would also not be established. A potential example here concerns the above-mentioned trials establishing a

correlation between homeopathic interventions and improved health outcomes; the effectiveness of such interventions has not been established, since the proposed mechanism underlying such effectiveness has not been established (Clarke *et al.* 2014: 344). It is in this way that attempting to establish the existence of a mechanism can help to determine whether a correlation is causal.

Let us now look at the problem of external validity. Suppose that observational studies have established the existence of a correlation between a medical intervention and some health outcome in a target population. Suppose also that an efficacy claim has been established by some ideal randomized trials, namely, that the intervention is a cause of health outcome in some study population. On the present account, this requires establishing in the study population both that there is a correlation between the intervention and the health outcome, and that there exists a mechanism in this population that can explain the extent of this correlation. One way to establish that the intervention is a cause of the health outcome in other target populations is by establishing that there also exists a sufficiently similar mechanism in the target population (Clarke *et al.* 2014: 347; Steel 2008). And failing to establish the existence of such a similar mechanism will lead to failing to establish that the intervention will cause the health outcome in a distinct target population, as in the benaxoprofen example above (Worrall 2007: 994-5). It is in this way that attempting to establish the existence of a sufficiently similar mechanism can help to determine the conditions under which randomized trials have external validity. (Williamson (2019) provides a more detailed account of the role of establishing mechanisms in determining the external validity of clinical studies.)

One worry is that this appeal to mechanisms reintroduces all the problems with mechanistic reasoning, namely, the problem of story-telling and the problem of incompleteness (cf. Solomon 2015: 116-24). However, the idea is that a causal claim about the effectiveness of a medical intervention should not be based only upon an established mechanism, but instead also upon an established correlation, where this correlation provides evidence that the intervention has a net effect on the health outcome, thus addressing the problem of incompleteness. Moreover, a causal claim is not established by being based upon a mere story about a mechanism, but instead by being based in part upon an established claim about a mechanism, where establishing such a claim requires evidence that may come from mechanistic studies into the details of the relevant mechanisms, that is, the relevant entities, activities, and their organization (Parkkinen *et al.* 2018: 14). (A list of examples of sources of evidence of mechanisms is given in Table 1 in Williamson (2019: 35).)

Now, it might be argued that there are cases in which mechanistic studies were not required to establish a causal claim about a medical intervention. Indeed, Howick objects that ‘there are many counterexamples where medical interventions have been [rightly] accepted on the basis of evidence from comparative clinical studies alone’ (2011b: 930). But this objection seems to assume that the existence of a mechanism can be established only by mechanistic studies. However, Phyllis Illari has argued that ‘there is no principled distinction between the *kinds* of empirical work by which we get evidence of mechanisms, and evidence of [correlation]—although in practice for any particular case in the health sciences these different items of evidence are usually got from different studies’ (Illari 2011: 145). Indeed, it has been argued in particular that sufficiently many well-conducted randomized trials that establish a large enough correlation can themselves provide establish the existence of a mechanism, for instance, if there is no other plausible explanation of such an established

correlation (Williamson 2019: 43-5). So the claim that clinical studies are sometimes sufficient to establish a causal claim about a medical intervention need not go against the Russo-Williamson thesis. And it certainly does not go against the claim that mechanistic studies can be helpful in those cases where the available clinical studies fail to establish the effectiveness of a medical intervention. At least where such studies are available, the current best evidence often consists of a combination of mechanistic studies and clinical studies.

Another worry is that there is nothing novel about the claim that causal claims can be established by relying upon mechanistic studies as well as comparative clinical studies.

Indeed, EBM proponents acknowledged from the outset that:

The dearth of adequate evidence demands that clinical problem solving must rely on an understanding of pathophysiology. Moreover, a good understanding of pathophysiology is necessary for interpreting clinical observations and for appropriate interpretation of evidence (especially in deciding on its generalizability). (Guyatt *et al.* 1992: 2423)

Presumably, a good understanding of pathophysiology can ultimately be traced back, through a sound medical education, to a number of mechanistic studies. Moreover, a published paper describing the results of a clinical study on some medical intervention will typically include a discussion of the intervention's possible mechanism of action (cf. Parkkinen *et al.* 2018: 67). Arguably, there has therefore always been an acknowledged role for relying upon both comparative clinical studies and mechanistic studies. Indeed, perhaps a reliance also upon mechanistic studies is precisely required for the use of evidence from randomized trials to count as 'judicious' (cf. Sackett *et al.* 1996: 71).

But to count as evidence-based, the use of current best evidence is also supposed to be ‘conscientious’ and ‘explicit’ (cf. Sackett *et al.* 1996: 71). In other words, it requires ‘conducting an efficient search of the literature; selecting the best of the relevant studies and applying rules of evidence to determine their validity’ (Guyatt *et al.* 1992: 2420). EBM provides a wealth of guidance to ensure the conscientious and explicit use of evidence from clinical studies, for example, by providing explicit guidance on how exactly to find and critically evaluate relevant clinical studies in an efficient, transparent, and reproducible manner (cf. Straus *et al.* 2019: 124). However, little guidance is provided on how exactly to find and critically evaluate mechanistic studies in a similar manner. So any use of mechanistic studies is therefore arguably neither explicit nor conscientious. The EBM+ approach agrees that evidence from comparative clinical studies should be used conscientiously, explicitly, and judiciously, but also stresses the conscientious, explicit, and judicious use of evidence from mechanistic studies (cf. Parkkinen *et al.* 2018: 3-4). In their words, ‘if EBM was a useful first approximation to evidence evaluation, then EBM+ is intended as a second, improved, approximation’ (Parkkinen *et al.* 2018: 7). In particular, the EBM+ approach provides an explicit procedure for evaluating evidence from mechanistic studies alongside evidence from clinical studies.

Broadly speaking, the EBM+ procedure is as follows. (A pictorial representation of this procedure in broad outline is captured by Figure 1.) The first step is to find and evaluate the available clinical studies in accordance with the guidance provided by EBM. (That is, we start in the bottom left-hand corner of Figure 1.) This will result in a body of evidence that can then be ranked in terms of its overall quality: high, moderate, low, or very low quality (Parkkinen *et al.* 2018: 26). The quality of this evidence in part determines the status of two

different claims: on the one hand, a *correlation claim*, namely, that the intervention is appropriately correlated with some health outcome in the study population; on the other hand, a *general mechanistic claim*, namely, that there exists a relevant mechanism in the study population that can explain the extent of this correlation. (That is, we move along the left-hand arrows of Figure 1.) In particular, it partly determines whether each claim is either: (i) established; (ii) provisionally established; (iii) arguably true; (iv) speculative; (v) arguably false; (vi) provisionally ruled out; or (vii) ruled out (Parkkinen *et al.* 2018: 27). For example, a claim is *established* when high quality evidence warrants a high level of confidence in that claim; a claim is merely *provisionally established* when it is instead moderate quality evidence that warrants a high level of confidence in that claim.

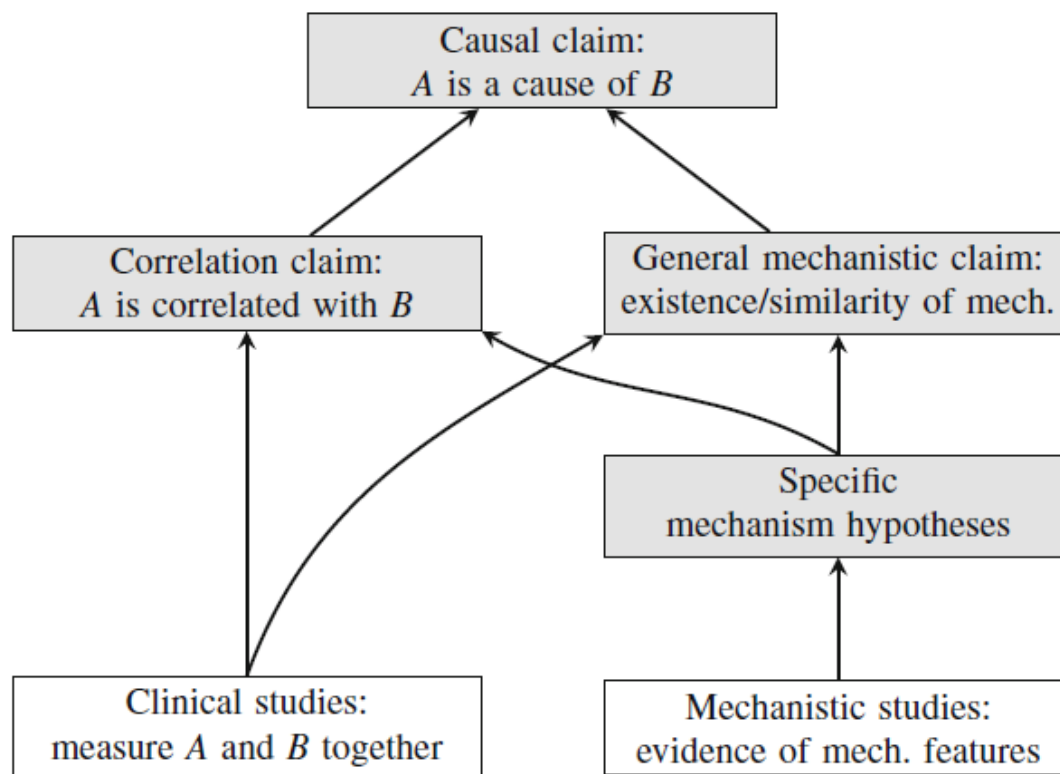


Figure 1: How to evaluate evidence according to the EBM+ approach (Parkkinen *et al.* 2018: 28).

In turn, the statuses of the correlation claim and the general mechanistic claim help to determine the status of *a causal claim*, namely, the claim that the medical intervention causes the health outcome in the study population. (That is, we move to the causal claim box from the correlation claim box and the general mechanistic claim box in Figure 1.) In particular, the status of this causal claim is the same as the lower of the statuses of the correlation claim and the general mechanistic claim (Parkkinen *et al.* 2018: 92). For instance, if the correlation claim is established but the general mechanistic claim is merely provisionally established, then the causal claim is likewise merely provisionally established. Now, it may be that the evidence from ideal clinical studies is sufficient to give the status of established to both the correlation claim and the general mechanistic claim; such evidence may then also give the status of established to this causal claim. All of this is in line with the Russo-Williamson thesis. However, it may also be that the evidence from clinical studies alone is insufficient to give the status of established to the causal claim, due to the problem of non-causal correlations: the available evidence establishes the correlation claim, but it fails to establish the causal claim, since it fails to establish the general mechanistic claim. Perhaps the general mechanistic claim is merely provisionally established on the basis of the clinical studies.

The next step then is to find and evaluate also the evidence from mechanistic studies (Parkkinen *et al.* 2018: 63-90). (That is, we carry on to the bottom right-hand corner of Figure 1.) Together with the first step, this will result in a total body of evidence consisting of both clinical studies and mechanistic studies. This total body of evidence can then also be ranked in terms of its overall quality, and then relied upon to determine the new statuses of both the correlation claim and the general mechanistic claim. (That is, we move along the right-hand arrows of Figure 1.) Of course, mechanistic studies are not typically most concerned with testing a general mechanistic claim; rather, they are more concerned with

testing a particular or *specific mechanism hypothesis* by investigating the proposed entities and activities. But such mechanistic studies can nevertheless still provide evidence that helps to determine the status of the logically weaker general mechanistic claim, that there simply exists a mechanism, without specifying its particular details. (That is, we move from the specific mechanism hypotheses box to the general mechanistic claim box in Figure 1.) And if this total evidence determines that both the correlation claim and the general mechanistic claim are established, then it also determines that the causal claim is now established, since the status of the causal claim is the same as the lower of the statuses of the correlation claim and the general mechanistic claim. (That is, we move to the causal claim box from the correlation claim box and the general mechanistic claim box in Figure 1.) In other words, *the efficacy claim* is now established, that is, it is now established that the medical intervention causes the health outcome in the study population. However, this does not mean that the combined clinical studies and mechanistic studies have thereby established the distinct *effectiveness claim*, namely, that the medical intervention causes the health outcome in some target population. Perhaps this causal claim is only provisionally established. This is just the old problem of external validity (Cartwright 2007, 2010).

The final step then is to determine the status of another general mechanistic claim, namely, the claim that there exist similar mechanisms in the target population as in the study population. Again, this status is determined in part by the quality of the total body of evidence, including clinical studies and mechanistic studies. And the status of this general mechanistic claim can be used to determine the status of the effectiveness claims that the medical intervention causes the health outcome in the target population. In particular, if the efficacy claim is provisionally established on the basis of the total evidence, but it is also established that there are sufficiently similar mechanisms at work in the target population,

then the effectiveness claim is also established. (Other combinations of statuses and their impact on an effectiveness claim are covered in Parkkinen *et al.* (2018: 94-7).)

I have here described the procedure only in broad outline. Parkkinen *et al.* (2018) provides the more detailed account. It also provides guidance on how exactly to carry out each step of this procedure in order to help ensure that the use of the total body of evidence remains conscientious, explicit, and judicious. For example, guidance is given on how to effectively gather and critically evaluate mechanistic studies alongside clinical studies (2018: 63-98). And Parkkinen *et al.* describe exactly how to determine the status of both a correlation claim and a general mechanistic claim on the basis of a body of evidence that includes both clinical studies and mechanistic studies, for instance, by giving guidance on how to evaluate: (i) the methods used in mechanistic studies; (ii) the implementation of these methods; (iii) the results of the mechanistic studies (2018: 80-2). Moreover, tools are provided to facilitate with these tasks (Parkkinen *et al.* 2018: 37-59).

The EBM+ approach thus provides procedural guidance to ensure the conscientious, explicit, and judicious use of the current best evidence, where this current best evidence includes evidence from mechanistic studies as well as comparative clinical studies. Of course, this approach is not completely uncontroversial. One important worry is that it fails to adequately address the biasing effects of financial or other conflicts of interest in medicine (Andreoletti and Teira 2019; Holman 2019; Howick 2019). In response, Williamson argues that taking on board mechanistic studies as well as clinical studies lowers the risk of such biasing effects by giving ‘less scope for any malleability with respect to individual judgements to influence the final assessment of causality’ (Williamson 2021: 206). In effect, the idea is that a combination of mechanistic studies and clinical studies helps to keep such biases in check,

since the mechanistic studies act as independent witnesses for the clinical studies, and vice versa.

Regardless, the EBM+ approach was never intended to be the final word on the matter. This point is acknowledged by Adam La Caze:

The next step is implementing the framework and the tools it provides into decision-making in medicine, public health, and policy. Opportunities for improving the tools and the evaluation framework will come from widespread implementation in range of contexts (La Caze 2019: 2).

In particular, it is by implementing the EBM+ approach that we may be able to determine whether the approach is more or less susceptible to the biasing effects of conflicts of interest than EBM. Indeed, Andreoletti and Teira suggest comparing approaches against appropriate empirical benchmarks in a pilot committee (2019: 1109). By doing this, we may also find ways to address any such biasing effects, thereby improving our methods for establishing the effectiveness of medical interventions.

Conclusion

In this paper, I have provided an introduction to some recent work in the philosophy of medicine. I have also drawn out the implications of this work for the practice of establishing causal claims about the effectiveness of medical interventions. In particular, this recent work in the philosophy of medicine argues that such causal claims are best established by evaluating evidence from mechanistic studies as well as clinical studies. And this work has

led to guidance on how to evaluate such a diverse body of evidence in a conscientious, explicit, and judicious manner, namely, the EBM+ approach. Although this paper has focused on guidance concerning treatment decisions, similar guidance can be provided for other medical decisions, for example, deciding whether an exposure is a cause of disease (Parkkinen *et al.* 2018: 101-10).

Acknowledgements

I would like to acknowledge very helpful comments from David Corfield, Isabelle Drouet, Erica Moodie, Yafeng Shan, and Jon Williamson.

References.

Andreoletti, M. and D. Teira. 2019. Rules versus standards: what are the costs of epistemic norms in drug regulation? *Science, Technology, & Human Values*, 44(6): 1093-1114.

Ashcroft, R.E. 2004. Current epistemological problems in evidence based medicine. *J Med Ethics*, 30: 131-135.

Auker-Howlett, D. and M. Wilde. 2020. Reinforced reasoning in medicine. *Journal of Evaluation in Clinical Practice*, 26(2): 458-464.

Cartwright, N. 2007. Are RCTs the gold standard? *BioSocieties*, 2: 11-20.

Cartwright, N. 2010. What are randomized controlled trials good for? *Philosophical Studies*, 147: 59-70.

Cartwright, N. 2011a. A philosopher's view of the long road from RCTs to effectiveness. *The Lancet*, 377: 1400-1401.

Cartwright, N. 2011b. Author's reply. *The Lancet*, 378: 1697.

Clarke, B., D. Gillies, P. Illari, F. Russo, and J. Williamson. 2014. Mechanisms and the evidence hierarchy. *Topoi*, 33: 339-360.

Cucherat, M., M.C. Haugh, M. Gooch, and J.P. Boissel. 2000. Evidence of clinical efficacy of homeopathy. A meta-analysis of clinical trials. *European Journal of Clinical Pharmacology*, 56(1): 27-33.

Gillies, D. 2019. *Causality, Probability, and Medicine*. Abingdon, Oxon: Routledge.

Guyatt, G., J. Cairns, D. Churchill, et al. 1992. Evidence-based medicine: A new approach to teaching the practice of medicine. *JAMA*, 268(17): 2420-2425.

Guyatt, G.H, A.D. Oxman, G.E. Vist, R. Kunz, Y. Falck-Ytter, P. Alonso-Coello, and H.J. Schuenemann. 2008. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, 336: 924-6.

Hill, A.B. 1965. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58(5): 295-300.

Holman, B. 2019. Philosophers on drugs. *Synthese*, 196: 4363-4390.

Howick, J. 2011a. *The Philosophy of Evidence-Based Medicine*. Wiley-Blackwell.

Howick, J. 2011b. Exposing the vanities—and a qualified defense—of mechanistic reasoning in health care decision making. *Philosophy of Science*, 78(5): 926-940.

Howick, J. 2019. Exploring the asymmetrical relationship between the power of finance bias and evidence. *Perspectives in Biology and Medicine*, 62(1): 159-188.

Illari, P. 2011. Mechanistic evidence: Disambiguating the Russo-Williamson thesis.

International Studies in the Philosophy of Science, 25(2): 139-157.

Jerkert, J. 2021. On the meaning of medical evidence hierarchies. *Philosophy of Medicine*,

2(1): 1-21.

La Caze, A. 2019. Better evaluating mechanisms in medicine. *Journal of Evaluation in*

Clinical Practice, 25(6): 1228-1231.

Machamer, P. L. Darden, and C.F. Craver. 2000. Thinking about mechanisms. *Philosophy of*

Science, 67(1): 1-25.

Parkkinen, V-P., C. Wallmann. M. Wilde. B. Clarke, P. Illari, M.P. Kelly, C. Norell, F.

Russo, B. Shaw, and J. Williamson. 2018. *Evaluating Evidence of Mechanisms in Medicine:*

Principles and Procedures. Cham, Switzerland: Springer.

Petticrew, M. and I. Chalmers. 2011. Use of research evidence in practice. *The Lancet*, 378:

1696.

Russo, F., and J. Williamson. 2007. Interpreting causality in the health sciences. *International*

Studies in the Philosophy of Science, 21(2): 157-170.

Sackett, D.L., S.E. Straus, W.S. Richardson, W. Rosenburg, and R.B. Haynes. 2000.

Evidence-Based Medicine: How to Practice and Teach EBM (2nd ed.). Churchill Livingstone.

Sackett, D.L., W.M.C. Rosenberg, J.A.M. Gray, R.B. Haynes, and W.S. Richardson. 1996. Evidence based medicine: what it is and what it isn't. *BMJ: British Medical Journal*, 312(7023): 71-72.

Solomon, M. 2015. *Making Medical Knowledge*. Oxford: Oxford University Press.

Steel, D. 2008. *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford: Oxford University Press.

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

Strader, D.B., T. Wright, D.L. Thomas, L.B. Seeff. 2004. Diagnosis, management, and treatment of hepatitis C. *Hepatology*, 39(4): 1147-1171.

Straus, S.E., P. Glasziou, W.S. Richardson, and R.B. Haynes. 2019. *Evidence-Based Medicine: How to Practice and Teach EBM* (5th ed.). Elsevier.

Urbach, P. 1985. Randomisation and the design of experiments. *Philosophy of Science*, 52(2): 256-273.

Urbach, P. 1993. The value of randomization and control and clinical trials. *Statistics in Medicine*, 12: 1421-1431.

Wheeler, G. 2007. A review of the lottery paradox. In W. Harper and G. Wheeler (eds.), *Probability and Inference: Essays in Honour of Henry E. Kyburg, Jr.* London: College Publications, pp. 1-31.

Wilde, M. 2021. Mechanistic reasoning and the problem of masking. *Synthese*, 199: 6103-6118.

Williamson, J. 2019. Establishing causal claims in medicine. *International Studies in the Philosophy of Science*, 32(1): 33-61.

Williamson, J. 2021. The feasibility and malleability of EBM+. *Theoria*, 36(2): 191-209.

Worrall, J. 2002. What Evidence in Evidence-Based Medicine? *Philosophy of Science*, 69: S316-S330.

Worrall, J. 2007. Evidence in medicine and evidence-based medicine. *Philosophy Compass*, 2(6): 981-1022.