

# Research Gaps in the Philosophy of Evidence-Based Medicine

Alexander Mebius<sup>1</sup>, Ashley Graham Kennedy<sup>2</sup>, and Jeremy Howick<sup>1</sup>

<sup>1</sup>*Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK*

<sup>2</sup>*Wilkes Honors College, Florida Atlantic University, Boca Raton, Florida, US*

## Abstract

Increasing philosophical attention is being directed to the rapidly growing discipline of evidence-based medicine (EBM). Philosophical discussions of EBM, however, remain narrowly focused on randomization, mechanisms, and the sociology of EBM. Other aspects of EBM have been all but ignored, including: 1) the nature of clinical reasoning and the question of whether it can be standardized; 2) the application of EBM principles to the logic, value, and ethics of diagnosis and prognosis; 3) evidence synthesis (systematic reviews and meta-analyses); and 4) the nature and ethics of placebo controls. Philosophical analysis in each of these areas has the potential to enhance the discussion of EBM methodology and practice.

## 1 Introduction

Evidence-based medicine (EBM) was introduced less than 25 years ago and quickly became widely adopted in medical research and practice (Howick 2011a). Surgeons, general practitioners, and nurses are all required to engage in evidence-based practice, guidelines are supposedly evidence-based, and even policies are often touted as evidence-based. EBM has been defined as “the integration of best research evidence with clinical expertise and patient values” (Sackett et al. 1996: 71). Perhaps the most prominent feature of EBM is its view of what counts as good medical evidence. The

EBM view is that empirical findings from well-designed clinical studies provide the best guide to clinical practice. Such evidence—at least regarding intervention outcomes, again from the EBM perspective—often comes from randomized trials, and systematic reviews of such trials (Straus 2011).

EBM places a lower premium on evidence from observational studies because the design of these studies is not experimental, and is thus allegedly prone to bias. Evidence derived from clinical experience and preclinical or basic research is valued even lower than observational evidence (if at all). In practice, EBM proponents usually advise against treatment decisions based on knowledge of physiology (mechanisms) and expert opinion alone: “Physiologic studies and unsystematic clinical observation provide the weakest inference about treatment effects” (Guyatt et al. 2002: 7).

The strong claims EBM makes about what counts as good evidence, however, have been controversial from the outset (Tonelli 1998), and philosophers have paid increasing attention to these claims. As a result, the philosophical treatment of EBM has been predominantly critical and focused on a narrow range of issues. Philosophers of science have written extensively about: (i) the special role of randomized trials within EBM as evidence for therapeutic claims, (ii) the special role of mechanisms in evaluating and guiding treatment; and (iii) the sociological consequences of EBM (Howick, Kennedy, and Mebius 2015).

It is likely that philosophers have focused on this narrow range of issues for historical reasons. For example, Bayesian philosophers of science criticized the role of randomization within frequentist statistics prior to and independently of EBM, and Worrall’s seminal critique of randomized trials was inspired by Howson’s and

Urbach's prior work on Bayesian reasoning (see Urbach 1985; Howson and Urbach 2005). The focus on randomization could also be due to the way EBM has been implemented in practice. While the founders of EBM rejected the notion that evidence was restricted to randomized trials (Gilbert et al. 1996; Sackett et al. 2005), much EBM training, both in medical schools and in professional development courses, focuses on the critical appraisal of randomized trials. This (over)emphasis on randomized trials (and trials of therapeutic interventions generally) can be taken to imply a belief that randomized trials are the only important source of evidence in the practice of medicine, despite what other proponents of EBM have written.

Likewise, we suspect that the philosophical attention devoted to the evidential role of mechanisms within EBM system (Clarke et al. 2014), can be traced to groundbreaking work by Cartwright (1989), as well as to Machamer's, Darden's, and Craver's often cited paper, "Thinking About Mechanisms". Both Cartwright and Machamer et al. make a case for the value of mechanisms for solving many longstanding problems in both philosophy of science and beyond. These authors also explicitly draw on a rich philosophical tradition invested in questions about causal explanation (beginning with Hempel 1965 and Nagel 1961). Craver's (2007) famous treatise on mechanistic explanation, for example, draws explicitly on earlier philosophical research, such as Cummins' (1975) classical account of causal role functions and Salmon's (1984) theory of causality. Cummins, for instance, proposes that a causal analysis of a mechanism explains "how [the mechanism] is able to produce the product by appeal to certain capacities of the workers/machines and their organization into an assembly line" (760).

Further, the philosophical attention given to the sociological consequences of EBM is unsurprising because of EBM's sociological success. In just over two decades, EBM has established itself as the dominant paradigm in medical care. EBM coursework is a requirement for medical students and continuing professional development programs, and therefore dozens of textbooks, training manuals, and histories of EBM have been (or are being) produced. Thus philosophers have rightly turned their attention to an analysis of the impact this EBM emphasis has had in the medical community, as well as in the public and private sectors.

Recently, some philosophers have examined evidence-based policy (Cartwright and Hardie 2012) and the role of values in the EBM system (Fulford, Peile, and Carroll 2012). However, other potentially compelling characteristics of EBM, including the role of evidence synthesis, evidence for prognostic markers, and diagnostic and clinical reasoning in the EBM system, have been largely ignored, and the philosophical work on the roles of expertise, blinding, and placebo controls in EBM is present but sparse.

In what follows, we shall examine five of the research gaps in the philosophy of EBM: the nature of clinical reasoning (section 2), its application to the logic, value, and ethics of diagnosis (section 3) and to prognosis (section 4); the evidential roles of systematic reviews and meta-analysis (5); and the nature and ethics of placebo controls (section 6).

## **2 Clinical Reasoning**

Clinical expertise is frequently *mentioned* in the EBM literature. It is therefore somewhat paradoxical that there is very little detail or discussion of the role of clinical expertise in clinical reasoning. Following Sox et al. (2007) and Stempsey (2015), we take clinical reasoning to be the complex process of combining inductive, deductive, and abductive reasoning together with tacit knowledge and weighing various forms of evidence to facilitate patient diagnosis and care. Guidance on integrating clinical reasoning into evidence-based practice (EBP) has been tenuous at best. Importantly, the issue of clinical reasoning has been dealt with extensively in the philosophy of medicine literature (e.g., Whitbeck 1981; Upshur 1997; Moreira et al. 2008; Edmondson et al. 2009; Braude 2012a; 2012b; Marcum 2013), but the relative value of clinical reasoning and methods for its integration into EBM have largely been ignored by philosophers. Here, we suggest a number of conceptual and practical questions in this area that we believe could benefit from philosophical analysis and potentially guide the development of a conceptual framework for interpreting and integrating clinical reasoning into EBM.

First, there is the question of how clinical reasoning should be understood and employed in the specific medical domains where EBM methodology and principles are applied; most existing discussion has rather focused on the role of clinical reasoning in science or medicine in general. Similarly, the specific role expertise plays in different realms of medical research (e.g., diagnosis, prognosis, and treatment) could be clarified. Finally, the question of how specific forms of clinical reasoning, such as diagnostic and prognostic reasoning, ought to be positioned and practiced within EBM could also benefit from philosophical analysis. In the following sections, each of these identified research gaps will be addressed in more detail.

It has often been noted—not least by EBM’s proponents—that clinical reasoning has an important role to play in evidence-based healthcare (Straus et al. 2011). However, what the role of clinical reasoning is, and how the concept might be integrated as a feature in EBM has not been dealt with in sufficient detail. Take, for example, the question of how components of clinical reasoning are organized and understood, and how this relates to other components of EBM. Should we take clinical reasoning to be an integration of research evidence and personal clinical experience, or should we understand clinical reasoning as something else entirely? How does this fit with the EBM view of what counts as medical evidence, given that the opinion of experts is often taken to be the weakest form (Prasad and Ioannidis 2014)?

Also as yet unaddressed in the philosophical literature is the question of whether clinical experience should count as evidence of a lower form than comparative studies or whether, perhaps, it should not count as evidence at all within the EBM system. Proponents of EBM are not clear on this issue, and practicing clinicians often rely on their experience as a form of evidential support when making clinical decisions. The philosophical literature lacks a clear exposition of the relative epistemic role and credibility of clinical reasoning within the EBM framework.

Further, a key tenet of the EBM methodology is the assumption that medical practice should be standardized and taught. The reticence about clinical reasoning in EBM texts and curricula could be due to the fact that clinical reasoning is often characterized as using both quantitative and qualitative forms of evidence, and qualitative evidence is often presented in nonstandard forms. It could follow that

clinical reasoning is inherently resistant to standardization. However, whether there is a way to reliably standardize nonquantitative evidence and its use in EBP—and, if so, how—remains an open philosophical question. Moreover, even if qualitative evidence and its clinical use can be standardized, philosophers and clinicians must still ask whether such standardization would be beneficial for EBP or whether room should remain for nonstandardized forms of reasoning in the clinic.

Another philosophically interesting issue in this area is the question of whether clinical reasoning is a form of practical reasoning. Practical reasoning is concerned with action (Wallace 2014), and clinical reasoning in medicine certainly fits this description, because medicine is an applied practice. However, if clinical reasoning is a form of practical reasoning (Korsgaard 1984; Parfit 1997), the question arises of to what reasoning norms clinicians are subject, and whether and how practical reasoning and its norms can be standardized and taught in the context of EBM.

### **3 Diagnosis**

Before treatment or prognostic evaluation of a patient can begin, there must be a diagnosis. Without an accurate diagnosis, treatments are likely to fail, and prognoses are likely to be inaccurate. Therefore, it is somewhat surprising that there has been little discussion of evidence-based diagnosis in the general philosophy of medicine literature (e.g., Førde 1998; Weber and Provijn 1999; Lucchiari and Pravettoni 2012; Staden 2013). We believe this is a significant research gap, given the importance of reliable diagnoses for EBP in medicine.

Philosophical questions pertaining to diagnostic practice in the context of EBM can be roughly divided into those that concern diagnostic reasoning and those that

concern diagnostics (i.e., tests and procedures used in the process of medical diagnosis). Regarding reasoning, questions include: (a) whether there is a logic of diagnosis, and if so; (b) whether this logic is computable or generalizable (and thus teachable); and (c) whether diagnostic reasoning should be considered scientific, deontic, or both. Regarding diagnostics, questions concern the diagnostic tests and procedures themselves and how they should be evaluated and implemented. For example, one might ask whether information about sensitivity and specificity derived from diagnostic cohort studies is sufficient to analyze the accuracy of diagnostic tests, or whether randomized trials are needed instead (see Brozek et al. 2008 for an informative nonphilosophical treatment of these questions). Further, as we shall see, even after the accuracy of a given test has been established, the question of whether it is clinically valuable remains.

Discussions of the logic of diagnostic reasoning often begin by asking whether it resembles other types of scientific reasoning (Upshur 2005; Sadegh-Zadeh 2011). For example, does diagnostic reasoning follow the deductive-nomological model, is it case-based, is it inductive, or is it related to model-based reasoning in science and thus reliant on representation and/or similarity relationships? The latter approach, although briefly mentioned in the literature (Upshur and Colak 2005), is an underexplored area in philosophy of EBM. Similarly, while there is a large philosophy of science literature on the role of models and modeling in scientific practice, this literature has yet to be expanded into a philosophical discussion of medical models in evidence-based diagnosis. Another possibility is that the logic of diagnosis is some combination of the methods mentioned above, which might then relate to the ongoing philosophical controversy about whether medicine is an art



(Munson 1981), a science, or both (see Matthews for a historical overview of medical debates involving many of these issues).

If diagnostic reasoning can be described with a suitable logic, then it can be computed (Reiter 1987). There are currently many diagnostic programs, either available or in development, and dozens of studies have compared computer-aided diagnosis with expert diagnosis (Paul Meehl's famous 1954 book *Clinical Versus Statistical Prediction: A Theoretical Analysis and Review of the Evidence* was the first of these). Following the work of Meehl, philosophers of medicine have discussed important issues related to the role of computer-based diagnosis in general (e.g., Jeanpierre and Charpillet 2004), but have not dealt with these issues as they arise in the context of EBM in particular. However, whether such programs are helpful in EBP or could eventually replace human diagnosticians is under debate in the medical and computer science literature. These questions, in turn, stand to benefit from detailed philosophical analysis of several issues, such as, for instance, the question of whether computer-based algorithms can be helpful in determining what counts as evidence in the practice of clinical diagnosis in the context of EBM.

Related to issues concerning the logic and computability of diagnosis is the question of whether or not diagnostic reasoning should be understood as a branch of scientific reasoning, moral reasoning, or both. According to Sadegh-Zadeh (2011), diagnostic reasoning should be understood as both scientific and moral. But how does this view fit more generally within the EBM system, which is not specifically designed to address moral questions? Analysis of the interesting but problematic relationship between ethics and evidence in clinical diagnosis is warranted.

As we saw earlier, another important question related to diagnosis is how diagnostic tests should be evaluated. While randomized controlled trials (RCTs) have long been the EBM gold standard for determining treatment efficacy, diagnostic cohort studies have generally been accepted as sufficient for determining the accuracy of a diagnostic test (Sackett et al. 1996). However, the determination of accuracy alone cannot tell us whether a diagnostic test is clinically effective. In order to evaluate the clinical effectiveness of a given diagnostic, one needs to evaluate, on a given patient population, the effect of the test information *plus* any resulting treatment or prevention strategies. Some medical researchers (Bossuyt et al. 2006; Ferrante di Ruffano et al. 2012; Siontis et al. 2015) would insist that RCTs are the best way to do this. However, EBM—at least in its original formulation—claims that patient values and external evidence are equally important, and that evidence cannot dictate values in the medical context. This consequently leaves open the possibility that an accurate test for an untreatable and unpreventable disease might still have value for a patient. Philosophical analysis of what counts as value in a context would be helpful in understanding the role of diagnostic testing in EBP.

This debate over what counts as clinically valuable diagnostic testing relates to a number of more general philosophical themes about values and their epistemic role in scientific contexts (see, for instance, Douglas; Elliott and McKaughan 2014).

Hence, our understanding of diagnostic value in the context of EBM is related to the philosophical role of values in science generally. For example, we might decide that, in the context of EBM, the value of a diagnostic test is tied inextricably to patient

outcomes and is thus extrinsic. On the other hand, as suggested above, we might want to allow that a test that provides knowledge is clinically valuable even when it does not directly lead to improved patient health. Thus determining when an accurate diagnostic test is clinically valuable requires a working theory of medical value, which is currently lacking in the philosophy of medicine literature.

## **4 Prognosis**

After diagnosis, a patient might ask, “What’s going to happen to me?” Prognosis describes the probability that a patient with a specific clinical (or nonclinical) profile will develop a specific health outcome over a certain period of time. Prognostic information is essential for deciding when to start and when to stop treatment, and for monitoring disease progression. To make an informed choice about whether to start a treatment regimen, the patient and caregiver need to be able to compare the prognosis without the treatment with the prognosis with the treatment. Accurate prognoses of competing treatment options are also necessary for selecting the most effective treatment, as well as for choosing which individuals to include in a study.

Unfortunately, the reliability of available prognostic procedures remains limited (Riley et al. 2013).

Relatedly, the question of whether computer-aided prognosis could improve clinical reasoning and patient care remains to be addressed. The effects of error following human judgment are well documented in domains outside of medicine, such as aviation (Helmreich 2000). There is also a growing number of trials comparing clinical reasoning aided by real-time, computer-based, decision and action algorithms

with expert decision-making alone. The results from these trials so far suggest that computer aids may improve prognostic decisions in certain healthcare settings, such as trauma care (Fitzgerald et al. 2011). The question of what specific forms of prognostic reasoning may lead to better evidence-based decisions and assessments in clinical practice could be examined from the perspective of philosophical work on decision theory, for example by considering the application of techniques for assessing personal probabilities, utilities, and risks in the field of decision analysis. Similarly scholars might investigate the potential utility of computer simulation techniques for solving analytical intractability issues and drawing conclusions about empirical phenomena from simulation models (Reiss 2011) or of using these to forecast and predict (see, e.g., Grüne-Yanoff 2011 for an interesting approach to health policy). Philosophical discussions on the reliability and predictability of computer-implemented models are also closely related to the debate in EBM over which form of *diagnostic* reasoning is most accurate (see above).

In EBM, as we have seen, the gold standard for testing the efficacy of medical treatments is the randomized controlled trial (Howick 2011a; Hansson 2014). Yet EBM practitioners consider observational (particularly cohort) studies as the primary source of information for providing prognostic evidence and answering prognostic questions (Greenhalgh 2014). Observational studies allow for a longer follow-up period, which is useful for many types of prognostic information. However, there is tension between the “best” EBM evidence for prognostic markers and the many worrisome issues EBM textbooks and systems of evidence highlight about observational studies (Guyatt et al. 2008; Straus et al. 2011). Philosophers of medicine may wish to investigate this apparent contradiction between best evidence in

general and best evidence for prognostic information to see whether and how it might be resolved, and, if it cannot be resolved, what the alternative solutions might be.

Another epistemic issue regarding the validity of studies of prognostic tests (irrespective of study design) is their heavy reliance on biomarkers as predictors of patient-relevant factors such as mortality and morbidity. The National Institutes of Health Biomarkers Definitions Working Group defines a biomarker as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” (Atkinson 2001: 91). Biomarkers are used to indicate changes in specific cells, molecules, enzymes, or hormones that correlate with particular health states, such as disease progression. One problem with biomarkers is that they are surrogates for clinical outcomes, and the causal link between the surrogate and clinical outcome is often spurious. The theoretical (Howick 2011a) and practical (Greenhalgh 2014) problems with surrogates for causal factors are well known. The inadequacy of prognostic markers is also due to poor methodological quality of the preclinical research in which biomarkers are established (Hirst et al. 2014).

Further, discussion about the potential value of biomarkers is likely to inform the ongoing philosophical debate about the evidential value of mechanisms (Howick 2011b; La Caze 2011; Andersen 2012; Bluhm 2013; Clarke et al. 2014; Jerkert 2015). This is because biomarkers represent steps along the mechanistic pathway linking interventions with clinical outcomes.

Philosophers are giving increasing attention to clarifying how consideration of mechanisms may facilitate thinking about causality in general (Craver and Darden

2014; Mebius 2014a), and about causal claims in evidence in particular. Philosophers of science have criticized EBM's stance on mechanisms as allegedly downplaying the value of related evidence derived from basic and preclinical research (Clarke et al 2014). Mechanism-guided information from preclinical studies is indeed often used to suggest starting doses, methods of dose-escalation, and early phase clinical trial designs (Chevret 2006).

In theory, the mechanism-based approach appears to have a sound basis for informing drug development and early-phase human trials, but scientific justification for this apparent value is lacking. Further, the concerns recently highlighted by medical researchers and statisticians regarding the validity of basic science findings (Bustin 2014; Sena et al. 2014) have received little attention from philosophers. Many methodologically and epistemologically controversial issues, such as the poor transferability of results from animal models to the human clinical situation (Pound and Bracken 2014), may warrant philosophical discussion. Such discussion could include rethinking the epistemic significance and normative status of the ability of mechanistic models to predict (or reproduce) clinical benefit in humans. For example, recent empirical studies suggest that 75% to 90% of major mechanistic findings presented in high-impact biomedical journals are irreproducible (Begley and Ioannidis 2015) and thus presumably false. Meanwhile, the mechanistic models that can be reproduced in laboratory settings rarely translate to successful treatments in the clinic (van der Worp et al. 2010; Ioannidis et al. 2014). One reason for the poor quality of basic science and preclinical findings may be that mechanism-oriented research is generally unsophisticated in addressing evidence synthesis, which we turn to below.

## 5 Evidence Synthesis

Research accumulates over time, soon making it impossible for any single researcher to take all previous primary research into account. It would be impossible for a medical researcher to consider all of the hundreds of trials that are published every year. Yet if previous research is not taken into account, we risk wasting resources by repeating what has already been done. In the case of medical research, this is a serious potential problem because we might overlook previous treatments shown to have beneficial or harmful effects. Some form of evidence synthesis is therefore required in medicine, as in any discipline, and, indeed, is supported by the principle of total evidence (Carnap 1962). Another benefit of synthesizing evidence systematically is that it prevents researchers and businesses from “cherry picking” studies that support a desired conclusion. In medicine, the problem of cherry picking is particularly nefarious because drug companies are not required to publish results of trials that fail to show a positive effect of a drug, or show harms of a drug (Jefferson et al. 2014). and indeed supported by the principle of total evidence. This means that claims about the therapeutic effectiveness of many pharmaceuticals are not based on all relevant evidence, leading to mistaken estimates of benefit or harm. Another advantage of systematic reviews is that, whereas small studies in isolation cannot detect small but important effects, the synthesis of many small studies may be able to.

In 1979, Archie Cochrane criticized the medical profession for not having synthesized all of the relevant evidence available to clinicians. In his words: “It is surely a great criticism of our profession that we have not organized a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomized controlled trials [RCTs]” (Cochrane 1979: 9). His challenge instigated an international

collaboration to develop the *Oxford Database of Perinatal Trials* in the 1980s (Starr et al. 2009). The *Database* eventually led to the establishment in 1993 of the Cochrane Collaboration, an international organization that produces systematic reviews of randomized trials and observational studies. The Cochrane Database of Systematic Reviews produces evidence syntheses that are regarded as the most methodologically rigorous and free from financial conflicts of interest.

Potentially interesting confusion and controversy about systematic reviews abounds. To begin, there is conceptual confusion. Several terms are often used interchangeably to describe evidence synthesis within medical research: (1) evidence synthesis; (2) research synthesis; (3) meta-analysis; (4) systematic review; and (5) Cochrane review. Research synthesis and evidence synthesis are overarching terms used to describe any form of amalgamating evidence. Systematic review is a term used to describe a review of research or evidence that involves a transparent and explicit search strategy. Given the *prima facie* benefits of evidence synthesis, it is unsurprising that systematic reviews are generally regarded as the highest quality of evidence for supporting claims of therapeutic effectiveness (OCEBM 2011; Straus et al. 2011). Meta-analysis is a term introduced by Gene Glass in 1976 to describe statistical pooling of results (Clarke 2015). Systematic reviews may or may not contain meta-analysis (Howick and Mebius 2014). Despite the differences between the terms, meta-analysis and systematic reviews are often used interchangeably by medical researchers and philosophers alike (Stegenga 2011). Philosophers are well placed to help shed light on the conceptual confusion.

The methodology for conducting systematic reviews in general and Cochrane reviews in particular is continually evolving. The first edition of a book describing systematic review methodology was less than 100 pages long (Chalmers and Altman 1995). Only



6 years later, the second edition included nearly 500 pages (Egger et al. 2001), and the current version of the Cochrane Handbook for Systematic Reviews is too long to feasibly contain in a single printed volume (Higgins and Green 2011). The method for conducting systematic reviews entails: (a) specifying the question to be addressed by the systematic review; (b) defining eligibility criteria for which studies to include; (c) identifying all potentially eligible studies; (d) applying eligibility criteria in ways that limit bias; (e) assembling as large a proportion as possible of the relevant information from the chosen studies; (f) analysing this information, and, if appropriate and possible, using meta-analysis and a variety of analyses; and (g) preparing a structured report (Chandler et al. 2014).

Each step of this method is open to philosophical challenges. For example, the questions asked in systematic reviews are often narrow and not relevant to the increasingly multi-morbid population (Trikalinos et al. 2014).

Next, the choice of studies to include, while transparent by requirement, is to question. For example, Cochrane methodology encourages the use of randomized trials and tends to exclude observational studies because they and other nonrandomized studies are believed to suffer from methodological issues that may increase the risk of producing inflated results. As suggested in the Cochrane Handbook:

“The Cochrane Collaboration focuses particularly on systematic reviews of randomized trials because they are more likely to provide unbiased information than other study designs about the differential effects of alternative forms of health care...We strongly recommend that review authors should not make any attempt to combine evidence from randomized trials and NRS [non-randomized studies].” (Higgins and Green 2008, section 13)

Paradoxically, this choice was undermined by a Cochrane review, which showed that there is no average difference between randomized trials and observational studies (Anglemyer et al. 2014). Again, paradoxically, rigorous systematic reviews have revealed the serious problem that most trials, especially those with unfavorable (from the researchers' point of view) results, are not published (Dwan et al. 2008; c.f. Doshi et al. 2013). Unpublished trials rarely feature in systematic reviews, and therefore undermine the very rationale for evidence synthesis, which is to amalgamate all relevant evidence.

Finally, although many of the statistical techniques used in systematic reviews, such as hypothesis tests of differences between a treatment and a placebo, are the same as those used in other studies, other statistical techniques are novel. For example, funnel plots were introduced specifically in the context of systematic reviews to detect publication bias, several measures of intra-study variability are used for systematic reviews, and there are two competing statistical methods for weighing studies (the fixed effect model and the random effects model). To our knowledge, philosophers have not scrutinized these models or the circumstances under which they can be assumed to hold, all of which are potentially controversial (both from a Bayesian and frequentist perspective).

For example, a skewed funnel plot assists in the inference, from data showing that there is a disproportionate number of small studies with positive results, that there is publication bias. This is because, the argument goes, chance dictates that there must be some hidden small negative studies that were not published and do not show up in the funnel plot. However, this inference assumes that there is no other reason that smaller studies are more likely to have positive results. This assumption requires

scrutiny. Further, the choice of fixed effect or random effects models for weighing studies within systematic reviews is critically important, in that these models represent very different assumptions about the data and can lead to very different results. For this reason, there is controversy even within the medical community about which method to use (Schmidt et al. 2009). Last, general questions about the epistemic value of the robustness of, or concordance with, medical evidence could be brought to bear on systematic review methodology (e.g., Hey 2014).

## **6 Placebo Controls**

Many hold that, among randomized trials, which are considered gold standard evidence for the existence of therapeutic effects, those that are double blind and placebo controlled are the best (Schulz et al. 1995). By contrast, others have argued that there is no widely accepted definition for a placebo and no agreement on what effect (if any) placebos (whatever they turn out to be) have (Howick 2011a). Further, there is ongoing controversy regarding what counts as an adequate placebo control for complex treatments such as acupuncture and exercise (Howick 2011a; White et al. 2002). Finally, the failure to adequately define placebos has also contributed to the ongoing debates about whether placebos are ethical in clinical practice (Foddy 2002) and clinical trials (Senn 2003; Howick 2009, c.f. Holman 2015).

Placebos are often referred to as inert substances (Beecher 1955) or nonspecific treatments (Shapiro and Morris 1978). However, in many contexts, the inertness criterion can be difficult to fulfill. Applying a glycerin stick, for example, is pharmacologically inert (in that nothing is absorbed into the blood stream), but it could not be used as a placebo for chapped lips. Moreover, some substances that are

decidedly not inert are sometimes intentionally prescribed simply for the placebo effect. These include (regrettably) antibiotics for viral infections, sham surgery, and saline injections. Indeed, as Adolf Grünbaum (1986) pointed out, even the proverbial sugar or bread pill will prove far from inert in patients with, respectively, insulin-dependent diabetes or gluten intolerance.

Further, placebos can be specific; that is, various kinds of placebo analgesia (a) exist and (b) operate through the release of endorphins (“natural opiates”) into the bloodstream (Benedetti 2009). This type of effect could be considered just as specific that produced by (assuredly non-placebo) penicillin in killing a pneumococcus. The term “specific” is also sometimes used to denote “well-defined”, or “quantitatively precise”, which contradicts the first two senses. However, estimates of placebo effects (if we accept them) illustrate that these effects can be quantified in much the same way that non-placebo effects are quantified (Hróbjartsson and Gøtzsche 2010; Howick et al. 2013).

Because of these discrepancies, serious placebo researchers commonly claim that the placebo cannot be defined, and instead either opt for accepting any treatment described as a placebo in the context of a clinical trial as a placebo (Hróbjartsson and Gøtzsche 2001) or suggest deleting the term from our vocabulary altogether (Nunn 2009). Until philosophers of medicine bring their conceptual analysis to bear on the problem, it may be premature to eliminate the term.

Additionally, the relationship between the terms placebo, placebo control, and placebo effect require analysis. Philosophical investigation may reveal that merely defining the term placebo will suffice: a placebo control would then simply be a

placebo used in a clinical trial, and a placebo effect would be the effect of a placebo (or placebo control). Certainly, it is likely that the three terms are related; however, it could be that the three terms require distinct definitions. Philosophers are arguably well placed to find out.

The debate over whether placebos are ethical in clinical practice hinges on whether placebos require deception—lying to patients by telling them that the placebo was a real treatment (this question, of course, relates to the philosophical debate over what counts as a lie, e.g., Kagan 1988). Foddy (2009) has argued that deception is ethical provided that it helps patients. If he is correct (and this is open to debate) then placebos can be ethical even if they do require deception. However, deception may not be required for placebos to have their effects. Several randomized trials comparing open-label placebos (i.e., patients know the treatments are placebos) with no treatment controls found that placebos can work even if they are given non-deceptively (Aulas and Rosner 2003; Sandler and Bodfish 2008; Kaptchuk et al. 2010).

There is a parallel debate concerning the ethics of placebos in clinical trials. Arguments about the ethics of placebo-controlled trials—as opposed to active-controlled trials that compare experimental interventions with established therapies—rest on the assumption that placebos are different in kind from other treatments. This assumption may be unjustified (Howick 2009). Philosophers of medicine could contribute to all of these ongoing debates, which are relevant to both philosophy and the practice of evidence-based healthcare.

## Conclusion

Philosophical analysis of EBM has been largely critical and has focused on randomization and mechanisms. The narrow focus arguably has historical roots: debates about the epistemic value of randomization can be traced to (pre-EBM) Bayesian critiques of frequentist statistics, and the alleged epistemic importance of mechanisms can be traced to philosophical discussion of causality. Other features of the EBM, which we believe warrant philosophical discussion, have been all but ignored. Here, we highlighted several interesting conceptual problems (e.g., with the terms systematic review, meta-analysis, and placebo control), epistemological dilemmas (e.g., the validity of diagnostic tests and prognostic markers), and more general philosophical problems (e.g., the relationship between the art and science of diagnosis, and the logic of clinical reasoning), all of which could greatly benefit from philosophical analysis.

Further, several of the research areas we have identified illustrate the intimate relationship between ethics and evidence in the philosophy of EBM. For instance, one could ask the philosophical questions of whether placebos are ethical in clinical trials (or in clinical practice), whether it is ethical to conduct diagnostic tests that are not deemed medically valuable, or how best to determine prognostic outcomes in the absence of clinical trial data. An examination of these questions would allow for a better understanding of how the relationship between ethics and evidence impacts medical decision-making for clinicians and patients alike.

Progress in the research areas we have identified would be not only intrinsically interesting from a philosophical perspective but also potentially helpful in applying insights from philosophy to medical practice and decision making.

## Works Cited

- Andersen, Holly. 'Mechanisms: What are they Evidence for in Evidence-Based Medicine?' *Journal of Evaluation in Clinical Practice* 18 (2012): 992–999.
- Anglemeyer A., Horvath, H. T. and L. Bero. 'Healthcare Outcomes Assessed with Observational Study Designs Compared With Those Assessed in Randomized Trials.' *Cochrane Database of Systematic Reviews* 4 (2014): MR000034.
- Atkinson, Arthur J, et al. 'Biomarkers and Surrogate Endpoints: Preferred Definitions and Conceptual Framework.' *Clinical Pharmacology & Therapeutics* 69 (2001): 89–95.
- Aulas, Jean-Jacques, and Iulius Rosner. 'Efficacy of A Non Blind Placebo Prescription.' *L'Encéphale* 29 (2003): 68–71.
- Beecher, Henry K. 'The Powerful Placebo.' *Journal of the American Medical Association* 159 (1955): 1602–1606.
- Begley, C. G., and J. P. A. Ioannidis. 'Improving the Standard for Basic and Preclinical Research.' *Circulation Research* 116 (2015): 116–126.
- Benedetti, Fabrizio. *Placebo Effects: Understanding the Mechanisms in Health and Disease*. Oxford: Oxford UP, 2009.
- Bluhm, Robyn. 'Physiological Mechanisms and Epidemiological Research.' *Journal of Evaluation in Clinical Practice* 19 (2013): 422–426.
- Bossuyt, Patrick M. M., et al. 'Comparative Accuracy: Assessing New Tests Against Existing Diagnostic Pathways.' *British Medical Journal* 332 (2006): 1089–1092.

- , Jeroen Lijmer, and Ben Mol. ‘Randomised Comparisons of Medical Tests: Sometimes Invalid, Not Always Efficient.’ *Lancet* 356 (2009): 1844–1847.
- Bradley, Ben. ‘Extrinsic Value.’ *Philosophical Studies* 91 (1998): 109–126.
- . ‘Two Concepts of Intrinsic Value.’ *Ethical Theory and Moral Practice* 9 (2006): 111–130.
- Braude, Hillel D. ‘Conciliating Cognition and Consciousness: The Perceptual Foundations of Clinical Reasoning.’ *Journal of Evaluation in Clinical Practice* 18 (2012a): 945–950.
- . *Intuition in Medicine: A Philosophical Defense of Clinical Reasoning*. Chicago: Chicago UP, 2012b.
- Brozek, Jeffrey, et al. ‘Grading Quality of Evidence and Strength of Recommendations in Clinical Practice Guidelines: Part 2 of 3. The GRADE Approach to Grading Quality of Evidence About Diagnostic Tests and Strategies.’ *Allergy* 64 (2009): 1109–16.
- Busting, S. A. ‘The Reproducibility of Biomedical Research: Sleepers Awake!’ *Biomolecular Detection and Quantification* 2: 35–42.
- Carnap, Rudolf. *Logical Foundations of Probability*. 2<sup>nd</sup> ed. Chicago: University of Chicago Press, 1962.
- Cartwright, Nancy. *Nature's Capacities and their Measurement*. Oxford: Clarendon, 1989.
- , and Jeremy Hardie. *Evidence Based Policy: A Practical Guide to Doing it Better*. Oxford: Oxford UP, 2012.
- Chalmers Iain, and Altman Douglas G. *Systematic Reviews*. London: BMJ Publications, 1995.



- Chandler J., et al. 'Methodological Standards for the Conduct of New Cochrane Intervention Reviews.' The Cochrane Collaboration. 11 August 2015.  
<[http://www.editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/MECIR\\_conduct\\_standards%202.1.pdf](http://www.editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/MECIR_conduct_standards%202.1.pdf)>.
- Clarke, Brendan, et al. 'Mechanisms and the Evidence Hierarchy.' *Topoi* 33 (2014): 339–360.
- Clarke, Mike. 'History of Evidence Synthesis to Assess Treatment Effects: Personal Reflections on Something That is Very Much Alive.' *JLL Bulletin: Commentaries On the History of Treatment Evaluation*. 3 October 2015.  
<<http://www.jameslindlibrary.org/articles/history-of-evidence-synthesis-to-assess-treatment-effects-personal-reflections-on-something-that-is-very-much-alive/>>.
- Chevret, S. *Statistical Methods for Dose-Finding Experiments*. Chichester: John Wiley & Sons, 2006.
- Cochrane, Archie L. '1931-1971: A Critical Review, With Particular Reference to the Medical Profession.' *Medicines for the Year 2000*. Eds. George Teeling-Smith and Nicholas Wells. London: Office of Health Economics, 1979. 1–11.
- Craver, C. F. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford UP, 2007.
- , L. Darden. *In Search of Mechanisms*. Chicago: University of Chicago Press, 2013.
- Cummins, Robert, 'Functional Analysis.' *Journal of Philosophy* 72 (1975): 741–765.
- Doshi, P., et al., 'Restoring Invisible and Abandoned Trials: A Call for People to Publish the Findings.' *BMJ (Online)* 346 (2013): f2865.

- Douglas, Heather. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press, 2009.
- Dwan K., et al. 'Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias.' *PLoS One* 3 (2008): e3081.
- Edmondson, Ricca, Pearce, Jane, and Markus H. Woerner. 'Wisdom in Clinical Reasoning and Medical Practice.' *Theoretical Medicine and Bioethics* 30 (2009): 231–247.
- Egger M., et al. 'Rationale, Potentials, and Promise of Systematic Reviews.' *Systematic Reviews in Health Care: Meta-Analysis in Context*. Eds. Egger et al. London: BMJ Books, 2001. 3–19.
- Elliott, K. C., and Daniel J. McKaughan. 'Non-Epistemic Values and the Multiple Goals of Science,' *Philosophy of Science* 81 (2014): 1–21.
- Ferrante di Ruffano, Lavinia, et al. 'Assessing the Value of Diagnostic Tests: A Framework for Designing and Evaluating Trials.' *British Medical Journal* 344 (2012): e686.
- Fitzgerald, Mark, et al. 'Trauma Resuscitation Errors and Computer-Assisted Decision Support.' *Archives of Surgery* 146 (2011): 218–225.
- Foddy, Bennett. 'A Duty to Deceive: Placebos in Clinical Practice.' *American Journal of Bioethics* 9 (2009): 4–12.
- Fulford, William, Ed Peile and Heidi Carroll. *Essential Values-Based Practice: Clinical Stories Linking Science with People*. Cambridge: Cambridge UP, 2012.
- Førde, Reidun. 'Competing Conceptions of Diagnostic reasoning—Is There a Way Out?' *Theoretical Medicine and Bioethics* 19 (1998): 59–72.

- Gilbert, Ruth, et al. 'Infant Sleeping Position and the Sudden Infant Death Syndrome: Systematic Review of Observational Studies and Historical Review of Recommendations from 1940 to 2002.' *International Journal of Epidemiology* 34 (2005): 874–887.
- Greenhalgh, Trish. *How to Read a Paper: The Basics of Evidence-Based Medicine*. 5<sup>th</sup> ed. BMJ Books, 2014.
- Grünbaum, Adolf. 'The Placebo Concept in Medicine and Psychiatry.' *Psychological Medicine* 16 (1986): 19–38.
- Grüne-Yanoff, Till. 'Agent-Based Models As Policy Decision Tools: The Case of Smallpox Vaccination.' *Simulation and Gaming: An Interdisciplinary Journal* 42 (2011): 219–236.
- Guyatt, Gordon H., et al. 'GRADE: An Emerging Consensus On Rating Quality of Evidence and Strength of Recommendations.' *British Medical Journal* 336 (2008): 924–926.
- , and D. Rennie (eds.) *User's Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago: American Medical Association, 2002.
- Hansson, Sven Ove. 'Why and for What Are Clinical Trials the Gold Standard?' *Scandinavian Journal of Public Health* 42 (2014): 41–48.
- Helmreich, Robert L. 'On Error Management: Lessons From Aviation.' *British Medical Journal* 320 (2000): 781–85.
- Hempel, C. G. *Aspects of Scientific Explanation*. New York: The Free Press, 1965.
- Hirst, Jennifer A., et al. 'The Need for Randomization in Animal Trials: An Overview of Systematic Reviews.' *PLoS ONE* 9 (2014): e98856.

- Higgins, J. P. T. and S. Green. *Cochrane Handbook for Systematic Reviews of Interventions*. 501<sup>st</sup> ed. Updated March 2011. The Cochrane Collaboration. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
- Hey, Spencer P. 'Robust and Discordant Evidence: Methodological Lessons from Clinical Research.' *Philosophy of Science* 82 (2014): 55–75.
- Holman, Bennett. 'Why Sugar Pills Are Not Placebos.' *Philosophy of Science* 82 (2015): 1330–1343.
- Howick, Jeremy, et al. 'Are Treatments More Effective Than Placebos? A Systematic Review and Meta-analysis.' *PLoS One* 8 (2013): e62599.
- . 'Exposing the Vanities—and A Qualified Defense—of Mechanistic Reasoning in Health Care Decision Making.' *Philosophy of Science* 75 (2011b): 926–940.
- and Alexander Mebius. 'In Search of Justification for the Unpredictability Paradox.' *Trials* 15 (2014): 480.
- , Kennedy, Ashley Graham and Alexander Mebius. 'Philosophy of Evidence Based Medicine.' *Oxford Bibliographies in Philosophy*. Ed. Duncan Pritchard. Oxford UP, 2015.
- . *The Philosophy of Evidence-Based Medicine*. Oxford: Wiley-Blackwell, 2011a.
- . 'Questioning the Methodologic Superiority of 'Placebo' Over 'Active' Controlled Trials.' *American Journal of Bioethics* 9 (2009): 34–48.
- Howson, Colin and Peter Urbach. *Scientific Reasoning: The Bayesian Approach*. 3<sup>rd</sup> ed. Chicago: Open Court, 2005.

- Hróbjartsson, Asbjørn and Peter C. Gøtzsche. 'Is the Placebo Powerless? An Analysis of Clinical Trials Comparing Placebo With No Treatment.' *New England Journal of Medicine* 344 (2001): 1594–1602.
- Hróbjartsson, Asbjørn and Peter C. Gøtzsche. 'Placebo Interventions for All Clinical Conditions.' *Cochrane Database of Systematic Reviews* 1 (2010): CD003974.
- Ioannidis, J. P. A., et al. 'Increasing Value and Reducing Waste in Research Design, Conduct, and Analysis.' *Lancet* 383 (2014): 166–175.
- Jeanpierre, Laurent and Francois Charpillet. 'Automated Medical Diagnosis With Fuzzy Stochastic Models: Monitoring Chronic Diseases.' *Acta Biotheoretica* 52 (2004): 291–311.
- Jefferson, T., et al. 'Neuraminidase Inhibitors for Preventing and Treating Influenza in Healthy Adults and Children.' *Cochrane Database of Systematic Reviews* 4 (2014): CD008965.
- Jerkert, Jesper. 'Negative Mechanistic Reasoning in Medical Intervention Assessment.' *Theoretical Medicine and Bioethics* 36 (2015): 425–437.
- Kaptchuk, T. J., et al. 'Placebos Without Deception: A Randomized Controlled Trial in Irritable Bowel Syndrome.' *PLoS One* 5 (2010): e15591.
- Kagan, Shelly. *Normative Ethics*. Boulder: Westview Press, 1988.
- Kennedy, Ashley Graham. 'Differential Diagnosis and the Suspension of Judgment.' *Journal of Medicine and Philosophy* 38 (2013): 487–500.
- Korsgaard, Christine. 'The Normativity of Instrumental Reason.' *Ethics and Practical Reason*. Eds. G. Cullity and B. Gaut, Oxford: Clarendon Press, 1997. 215–254.
- La Caze, Adam. 'The Role of Basic Science in Evidence-Based Medicine.' *Biology and Philosophy* 26 (2011): 81–98.

- Lucchiari, Claudio and Pravettoni, Gabriella. 'Cognitive Balanced Model: A Conceptual Scheme of Diagnostic Decision Making.' *Journal of Evaluation in Clinical Practice* 18 (2012): 82–88.
- Machamer, Peter, et al. 'Thinking About Mechanisms.' *Philosophy of Science* 60 (2000): 1–25.
- Marcum, J. A. 'The Role of Emotions in Clinical Reasoning and Decision Making.' *Journal of Medicine and Philosophy* 38 (2013): 501–519.
- Matthews, J. R. *Quantification and the Quest for Medical Certainty*. Princeton: Princeton UP, 1995.
- Mebius, Alexander. 'A Weakened Mechanism is Still a Mechanism: On the Causal Role of Absences in Mechanistic Explanation.' *Studies in History and Philosophy of Biological and Biomedical Sciences* 45 (2014a): 43–48.
- . 'Corroborating Evidence-Based Medicine.' *Journal of Evaluation in Clinical Practice* 20 (2014b): 915–20.
- Meehl, Paul E. *Clinical Versus Statistical Prediction: A Theoretical Analysis and A Review of the Evidence*. Minneapolis: University of Minnesota Press, 1954.
- Moons, Carl K. G. M., et al. 'Prognosis and Prognostic Research: What, Why, and How?' *British Medical Journal* 338 (2009): 1317–1320.
- Moore, 'Patrolling the Borders of Consequentialist Justifications: The Scope of Agent-Relative Obligations,' *Law and Philosophy* 27 (2008): 35–96.
- Moreira, Juan, et al. 'Bayesian Clinical Reasoning: Does Intuitive Estimation of Likelihood Ratios on an Ordinal Scale Outperform Estimation of Sensitivities and Specificities?' *Journal of Evaluation in Clinical Practice* 14: 934–940.
- Munson, Ronald. 'Why Medicine Cannot Be A Science.' *Journal of Medicine and Philosophy* 6 (1981): 183–208.

- Nagel, Ernest. *The Structure of Science*. New York: Harcourt Brace Jovanovich, 1961.
- Nunn, Robin. 'It's Time to Put the Placebo Out of its Misery.' *British Medical Journal* 338 (2009): b1568.
- OCEBM Levels of Evidence Working Group. 'The Oxford Levels of Evidence 2'.  
Oxford Centre for Evidence Based Medicine. 18 July 2015.  
<<http://www.cebm.net/index.aspx?o=5653>>.
- Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon Press, 1984.
- Pound, Pandora, and Michael B. Bracken. 'Is Animal Research Sufficiently Evidence Based to be a Cornerstone of Biomedical Research?' *British Medical Journal* 348 (2014): g3387.
- Prasad, Vinay and John P. A. Ioannidis. 'Evidence-Based De-Implementation for Contradicted, Unproven, and Aspiring Healthcare Practices.' *Implementation Science* 9 (2014): 1.
- Reiss, Julian. 'A Plea for (Good) Simulations: Nudging Economics Toward an Experimental Science.' *Simulation & Gaming* 42 (2011): 243–264.
- Reiter, Raymond. 'A Theory of Diagnosis from First Principles.' *Artificial Intelligence* 32 (1987): 57–95.
- Riley, R. D., et al. 'Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research.' *PLoS Medicine* 10 (2013): e1001380.
- Sackett, David, et al. 'Evidence Based Medicine: What It Is and What It Isn't.' *British Medical Journal* 13 (1996): 71–72.
- Sadegh-Zadeh, Kazem. 'The Logic of Diagnosis.' *Handbook of Philosophy of Science Vol 16: Philosophy of Medicine*. Ed. Fred Gifford. Elsevier BV, 2011. 357–424.

- Sandler, Adrian D., and James W. Bodfish (2008). 'Open-Label Use of Placebos in the Treatment of ADHD: A Pilot Study.' *Child: Care, Health and Development* 34 (2008): 104–110.
- Schmidt, Frank L., Oh, In-Sue, and Theodore L. Hayes. 'Fixed- Versus Random- Effects Models in Meta-analysis: Model Properties and an Empirical Comparison of Differences in Results. *British Journal of Mathematical and Statistical Psychology* 62 (2009): 97–128.
- Schulz, K. F., et al. 'Empirical Evidence of Bias. Dimensions of Methodological Quality Associated With Estimates of Treatment Effects in Controlled Trials.' *Journal of the American Medical Association* 273 (1995): 408–412.
- Senn, Stephen. *Dicing with Death*. Cambridge: Cambridge UP, 2003.
- Shapiro, Arthur, and Louis A. Morris. 'The Placebo Effect in Medical and Psychological Therapies.' *Handbook of Psychotherapy and Behavioural Change: An Empirical Analysis*. Eds. Sol Louis Garfield and Allen E. Bergin. New York: John Wiley & Sons, 1978. 369–410.
- Siontis, Konstantinos C., et al. 'Patient Outcomes in Randomized Comparisons of Diagnostic Tests: Still the Ultimate Judge: Reply to Letter by Ferrante di Ruffano and Deeks 'Test-treatment RCTs—Sheep in Wolves' Clothing.' *Journal of Clinical Epidemiology* (2015). DOI: 10.1016/j.jclinepi.2015.06.012.
- Sox, H. C., et al., *Medical Decision Making*. American College of Physicians, 2007.
- Staden, C. W. 'Examining Fallacies in Diagnostic Reasoning.' *Journal of Evaluation in Clinical Practice* 19 (2013): 528–530.
- Starr, Mark, et al. 'The Origins, Evolution, and Future of the Cochrane Database of Systematic Reviews.' *International Journal of Technology Assessment in Health Care* 25 (2009): 182–195.



- Stegenga, Jacob. 'Is Meta-analysis the Platinum Standard of Evidence?' *Studies in History and Philosophy of Biological and Biomedical Sciences* 42 (2011): 497–507.
- Stempsey, William. 'Applying Medical Knowledge: Diagnosing Disease.' *Handbook of the Philosophy of Medicine* (2015): 1–13.
- Straus, Sharon E., et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. 4<sup>th</sup> ed. Edinburgh: Churchill Livingstone, 2011.
- Tonelli, M. R. 'The Philosophical Limits of Evidence-Based Medicine. *Academic Medicine* 73 (1998): 1234–1240.
- Treadwell, Jonathan R., et al. 'A Framework for "Best Evidence" Approaches in Systematic Reviews.' *Methods Research Report*. (Prepared by the ECRI Institute Evidence-Based Practice Center under Contract No. HHSA 290-2007-10063-I.). Rockville: Agency for Healthcare Research and Quality, 2011.
- Trikalinos, T. A., Segal, J. B., and C. M. Boyd. 'Addressing Multimorbidity in Evidence Integration and Synthesis.' *Journal of General Internal Medicine* 29 (2014): 661–669.
- Upshur, Ross E. G., and Errol Colak. 'Argumentation and Evidence.' *Theoretical Medicine* 24 (2003): 283–299.
- . 'Certainty, Probability and Abduction: Why We Should Look to C.S. Peirce Rather Than Gödel for a Theory of Clinical Reasoning.' *Journal of Evaluation in Clinical Practice* 3 (1997): 201–206.
- . 'Looking for Rules in a World of Exceptions: Reflections on Evidence-Based Practice.' *Perspectives in Biology and Medicine* 48 (2005): 477–489.
- Urbach, Peter. 'Randomization and the Design of Experiments.' *Philosophy of Science* 52 (1985): 256–273.

- van der Worp, H. B., et al. 'Can Animal Models of Disease Reliably Inform Human Studies?' *PLoS Medicine* 7 (2010): e1000245.
- Wallace, R. Jay, 'Practical Reason', *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2014/entries/practical-reason/>.
- Weber, Erik and Dagmar Provijn. 'A Formal Analysis of Diagnosis and Diagnostic Reasoning'. *Logique et Analyse* 165 (1999): 61–180.
- Whitbeck, Caroline. What is Diagnosis? Some Critical Reflections. *Theoretical Medicine and Bioethics* 2 (1981): 319–329.
- White, Adrian J., et al. 'Clinical Trials of Acupuncture: Consensus for Optimal Treatment, Sham Controls and Blinding.' *Complementary Therapies in Medicine* 9 (2002): 237–245.
- Worrall, John. 'What Evidence in Evidence-Based Medicine?' *Philosophy of Science* 69 (2002): S316–S330.