Exposing the Vanities—and a Qualified Defense—of Mechanistic Reasoning in Health Care Decision Making
Author(s): Jeremy Howick
Reviewed work(s):
Source: *Philosophy of Science*, Vol. 78, No. 5 (December <sc>2011</sc>), pp. 926-940
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/10.1086/662561
Accessed: 28/06/2012 06:00

Exposing the Vanities—and a Qualified Defense—of Mechanistic Reasoning in Health Care Decision Making

Jeremy Howick[†‡]

Philosophers of science have insisted that evidence of underlying *mechanisms* is required to support claims about the effects of medical interventions. Yet evidence about mechanisms does not feature on dominant evidence-based medicine "hierarchies." After arguing that only inferences from mechanisms ("mechanistic reasoning")—not mechanisms themselves—count as evidence, I argue for a middle ground. Mechanistic reasoning is not required to establish causation when we have high-quality controlled studies; moreover, mechanistic reasoning is more problematic than has been assumed. Yet where the problems can be overcome, mechanistic reasoning can and should be used as evidence.

1. Introduction: Tension between Mechanistic Philosophy and Evidence-Based Medicine. Mechanisms are all the rage in current philosophical work on causality (Glennan 1996; Machamer, Darden, and Craver 2000; Bechtel and Abrahamsen 2005), where relatively strong ontological claims

†To contact the author, please write to: Centre for Evidence-Based Medicine, Oxford University, 23-38 Hythe Bridge Street, Oxford OX1 2ET, United Kingdom; e-mail: jeremy.howick@phc.ox.ac.uk.

‡This article is a development of work written in collaboration with Paul Glasziou and Jeffrey Aronson. It was written while I was a recipient of an MRC/ESRC Interdisciplinary Postdoctoral Fellowship (G0800055) and a lecturer in the Science and Technology Studies Department at University College London. Nancy Cartwright read and made invaluable comments on earlier drafts of this article, and John Worrall has been a continued inspiration. I am also grateful to Claire Blacklock, Brendan Clarke, Lindley Darden, Foad Dizadji-Bahmani, Donald Gillies, Stuart Glennan, Adam La Caze, Federica Russo, Jon Williamson, John Worrall, Raffaella Campaner, Maria Carla Galavotti, and Sir Iain Chalmers for useful comments and discussions. I also received enormously helpful comments from almost everyone at the conference Mechanisms and Causality in the Sciences, especially Alex Broadbent, Carl Craver, Phyllis McKay Illari, and Bert Leuridan.

Philosophy of Science, 78 (December 2011) pp. 926–940. 0031-8248/2011/7805-0018$10.00

are made on their behalf. Mechanisms are allegedly responsible for generating a good number of causal regularities of central interest in science. Expressing this view, Glennan states that "a mechanical theory of causation suggests that two events are causally connected when and only when there is a mechanism connecting them" (1996, 64). If mechanisms allegedly play such essential roles in underwriting causal regularities in the life sciences, it seems reasonable to expect that evidence about mechanisms should play a central role in supporting claims about the effects of medical interventions. These are, after all, just causal regularity claims with medical interventions as inputs and change in patient-relevant outcomes as outputs. Indeed, Russo and Williamson (2007, 159) insist on just this: "To establish causal claims, scientists need the mutual support of mechanisms and dependencies."

In stark contrast, evidence-based medicine (EBM) proponents do not rate mechanistic evidence at all highly. Mechanistic evidence does not even appear on the most recent, and arguably dominant, evidence-ranking scheme (Guyatt et al. 2008). Is this reasonable?

I will argue for a middle ground. Sometimes, evidence from mechanisms is not required, and in some cases comparative clinical studies are not required. Moreover, there are a number of overlooked problems that beset the use of mechanistic evidence. My overall conclusion, however, will not be in support of the EBM view that mechanistic reasoning is not evidence at all. Where problems with our mechanistic knowledge can be overcome, "mechanistic reasoning" can and should feature in EBM evidence ranking schemes. I will begin by defining my terms.

2. Terminology: Patient-Relevant Effects, Comparative Clinical Studies, Mechanisms, and Mechanistic Reasoning. I am concerned here with evidence that a health care intervention produces a patient-relevant outcome. Put simply, a patient-relevant outcome is one that makes people feel better or live longer as opposed to evidence about whether certain molecules latch onto a rat's cell receptor. Unless otherwise specified, I will be discussing evidence for patient-relevant effects.

Different types of evidence support claims that interventions are effective. The idea behind Mill's Methods (Mill 1843/1973), the "numerical" method (Louis 1836), the "statistical" method (Bernard 1957), and "difference-making" evidence (Russo and Williamson 2007) is similar because they all use what I will call "comparative clinical studies." Here, some (experimental) groups receive the experimental intervention, whereas other (control) groups do not. Then, if the outcomes differ significantly, the study counts as evidence that the intervention had an effect. In these studies the mechanism for how the intervention caused the outcome is mostly a "black box" (see fig. 1).

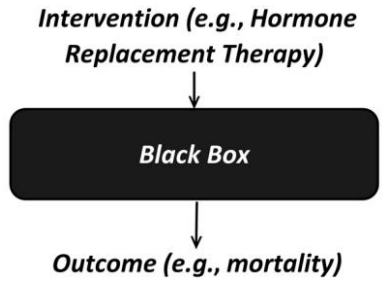

Figure 1. Comparative clinical studies.

For example, Petitti, Perlman, and Sidney (1987) compared records of 2,656 women who took hormone (estrogen) replacement therapy (HRT) with 3,437 who did not and followed them for 10 or more years to measure rates of coronary heart disease (CHD) and overall mortality. They found that HRT seemed to reduce mortality from all causes except cancer. Other studies yielded similar results (Stampfer and Colditz 1991), and HRT was widely prescribed.

Besides comparative clinical studies, evidence can be gained from mechanistic knowledge. A problem with exploring how mechanisms provide evidence is that "mechanism" has recently been characterized in several ways: "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (Machamer et al. 2000, 3). "A mechanism underlying a behavior is a complex system which produces that behavior by the interaction of a number of parts according to direct causal laws" (Glennan 1996, 52). "A nomological machine is a stable enough arrangement of components whose features acting in consort give rise to (relatively) stable input/output relations" (Cartwright 2009, 8). For present purposes these definitions are sufficiently similar. The heart (as a pump), the brain (as a "control center"), and the liver (as a detoxifying agent, among other things) are all mechanisms in the senses described above.

For example, the mechanisms involved in the apparent protective effect of HRT against CHD might be described as follows. First, the drug has to be metabolized. Mechanisms involved in getting orally administered drugs (such as estrogen) to their pharmacological targets (such as estrogen receptors on the cells) and out of the body are relatively well understood and referred to as ADME (mechanisms for absorption, distribution, metabolism, and excretion). Then estrogen binds on to estrogen receptors on various cells. In the mechanism expounded by Mendelsohn and Karas (1999), estrogens allegedly reduce blood lipid concentrations, cause va-

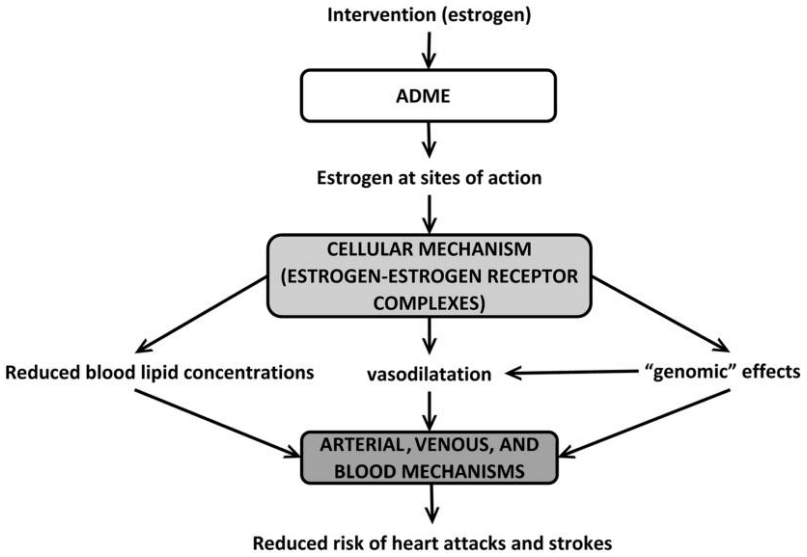

Figure 2. Simplified mechanistic reasoning for how hormone replacement therapy allegedly reduced cardiac events. ADME = absorption, distribution, metabolism excretion mechanisms.

sodilatation, and reduce the risk of blood clotting; this, in turn, reduces the risk of strokes, heart attacks, and pulmonary embolisms, all of which would otherwise lead to morbidity and mortality.

This example illustrates a central point of this article, namely, that descriptions of mechanisms alone—even if true descriptions—do not amount to evidence. Instead, evidence that mechanisms link an intervention with an outcome involves inferring from descriptions of all relevant mechanisms and knowledge of what happens to each mechanism under intervention to claims about the intervention's alleged benefits (see fig. 2). With that in mind, I will characterize evidence from mechanisms as follows:

> **Mechanistic reasoning** is an inferential chain (or web) linking the intervention (such as HRT) with a patient-relevant outcome, via relevant mechanisms.

It is generally possible to describe the mechanism at different levels. In the HRT example above, I might have distinguished between α and β estrogen receptors, or the action of these receptor complexes on different tissues (vascular and hepatic). The essential feature of mechanistic reasoning as I construe it is that the mechanisms provide the links in an

inferential chain (or web) connecting the intervention with the putative outcome.

3. Why We Do Not Require both Mechanistic Reasoning and Comparative Clinical Studies to Establish Causation. Russo and Williamson support their view that mechanistic reasoning is required alongside comparative clinical studies with "historical" and "theoretical" arguments. The historical argument is based on three anecdotes. First Semmelweis's comparative clinical study indicated that antiseptic procedures would reduce puerperal fever, but his findings were rejected (arguably) because the germ theory of disease was not available to explain why the procedure might work. Second, Doll and Hill's studies linking smoking and lung cancer were not fully accepted until the mechanism was established. Third, Warren and Marshall's claim that *Helicobacter pylori* caused peptic ulcers was practically laughed at until the mechanism was established.

Yet these same anecdotes could be used to point out the dangers with requiring mechanistic reasoning alongside comparative clinical studies to establish causation. Countless mothers and babies would have been saved had Semmelweis's intervention been adopted after the results from his comparative clinical study became available, and life expectancy in many countries would have risen decades earlier if government strictures on smoking had been introduced before the mechanism linking smoking and lung cancer was established. Hence, Gillies (whom Russo and Williamson ironically cite as a source) uses the same cases to argue that it is unwise to require mechanistic reasoning when there is strong evidence from comparative clinical studies (Gillies 2005, 180).

Moreover, there are many counterexamples where medical interventions have been accepted on the basis of evidence from comparative clinical studies alone. To name a few (see Howick 2011 for a more complete list), aspirin was used for a century before its analgesic mechanism was identified, the mechanism for general anaesthesia is still not well understood, and deep brain stimulation (DBS) is currently used to suppress tremors in patients with advanced Parkinson's and to cure other motor function disorders such as Tourette's Syndrome, yet researchers have not been able to identify the mechanism of DBS with any certainty.

The case against the view that the medical community requires both types of evidence can be made even more strongly. Sometimes evidence from tightly controlled comparative clinical studies is sufficient to overturn conflicting mechanistic reasoning (or other evidence). The apparent beneficial effects of antiarrhythmic drugs, rest for recovery, human growth hormone for hypercatabolism, and many other treatments had been supported by mechanistic reasoning, but tightly controlled comparative clinical studies subsequently suggested that the interventions in question were

useless or harmful for the same outcomes. Far from requiring both mechanistic reasoning and evidence from comparative clinical studies, the medical community often seems to regard high-quality comparative clinical studies as sufficient.

Russo and Williamson (2007) could insist that, for cases in which the mechanisms had not been identified, the medical community should have rejected claims that the treatments were effective, which brings us to their theoretical argument. Following Glennan, Russo and Williamson claim that "if there is no plausible mechanism from C to E, then any correlation is likely to be spurious" (2007, 159). It is true that comparative clinical studies sometimes support spurious relationships. We might find that there are more storks in areas where the birth rate is highest and mistakenly conclude that the storks caused the increase in birth rate.

But mechanistic reasoning does not prevent adoption of spurious hypotheses. Bloodletting was adopted on the basis of both mechanistic reasoning (derived from the humoural theory) and comparative clinical studies (observations of patients recovering after being bled), yet it is safe to say that bloodletting was useless or harmful in most cases. The example of bloodletting may be unfair because we now know that the evidence supporting its efficacy was weak. But this reinforces a point of this article: quality (tightly controlled, unbiased, etc.) evidence rather than quantity of evidence helps reduce the likelihood of spuriousness.

Moreover, there are more recent examples of less obviously flawed mechanistic reasoning. A more tightly controlled (randomized) trial suggested that women taking HRT were more likely to die from coronary heart disease, stroke, dementia, and breast cancer (Rossouw et al. 2002). This single trial sufficed to call the combined evidence from other ("observational") comparative clinical studies and mechanistic reasoning into question. Subsequently, it has been argued that the evidence from the different comparative clinical studies and mechanisms can be reconciled by the "timing hypothesis." The timing hypothesis is that the alleged protective benefit of HRT is limited to women age 50–59, whereas a larger proportion of the participants in the randomized trial were over 60. Even if we accept the timing hypothesis (and there are many who question it), the randomized trial sufficed to show that HRT is likely to be harmful for women over 60.

To be sure, it is circular to infer from cases where randomized trials overturned the results from other evidence to the claim that the randomized trial provided the "true" result. It is beyond the scope of this work to consider this objection in great detail (see Howick 2011). At the same time, at least in this case, there are independent reasons to believe the results of the randomized trial, at least as far as women above 60 are concerned. For example, the authors of the earlier studies observed that

mortality due to homicide was higher among women who did not take HRT. This (among other similar observations) implied that there were potentially confounding differences between women who chose (or were chosen by their doctors) to take HRT and those who did not. While many of these differences (such as smoking) are straightforward to control for, others are not. Randomization of large numbers of participants reduces the risk of these "baseline" confounders. There are good background reasons to accept results of the randomized trial over mechanistic reasoning. The methodology of comparative clinical studies has been rigorously studied for decades. While the methodology is far from perfect, the potential pitfalls with mechanistic reasoning I list in the next section have been all but ignored. (Then, in sec. 5 of this article, I take a step toward redressing the imbalance and propose desiderata for gauging the quality of mechanistic reasoning.)

I will now consider a weaker version of Russo and Williamson's argument—that a subset of hypotheses, namely, apparently implausible ones, requires support from both types of evidence. Consider the following example that appears to support such a view. In July 2000, 3,393 patients who had been admitted to hospitals between 1990 and 1996 were randomized to control and treatment groups. A remote, retroactive intercessory prayer was said for the well-being and full recovery of the intervention group (Leibovici 2001, 1450). The results indicated that although mortality was similar in intervention and control groups, "length of stay in hospital and duration of fever were significantly shorter in the intervention group than in the control group ($P = .01$ and $P = .04$, respectively)" (1450). The Leibovici study relies on a mechanism whereby effects precede causes. Any hypothesis that relies on such a mechanism can arguably be rejected out of hand. Here, mechanistic reasoning seems to guard against acceptance of apparently implausible hypotheses.

Yet there are at least two other strategies that can be used to rule out implausible hypotheses; therefore, mechanistic reasoning is not required. The first strategy is to appeal to the principle of total evidence. In the Leibovici example we might weigh the mechanistic reasoning with the evidence from comparative clinical studies and insist that the former outweighs the latter. The principle of total evidence does not, however, require that we use mechanistic reasoning. Instead, it demands that we weigh all relevant evidence and consider relative strengths rather than the quantity.

The second strategy is to evaluate the comparative clinical studies on their own grounds. Implausible hypotheses are either true or false. If true we would expect consistent detectable effects in unbiased comparative clinical studies. If such consistent effects are demonstrated, then we should recall the Semmelweis case and temper our skepticism regarding the plausibility of the hypothesis. On the other hand, if the implausible hypothesis

were false, we would not expect it to demonstrate consistent effects in unbiased comparative clinical studies; if not, we can refrain from accepting the hypothesis without appeal to any external (i.e., mechanistic) evidence. Leibovici's study falls in the latter category. The main outcome—mortality—was higher (although not statistically significantly) in the group that was prayed for. Next, it is unclear whether the positive outcomes (length of stay in hospital and duration of fever) were specified in advance. If we measure many outcomes, we are likely to find some that differ statistically significantly between groups due to chance alone. Finally, even if we accept the potentially cherry-picked secondary outcomes, their effects were minuscule. The median length of stay in the hospital was 8 days (interquartile range, 4–13) for the intervention group and 7 days (interquartile range, 4–16) for the control group. Small absolute differences must be interpreted with caution because they can be confounded by small, often undetected bias. I have not chosen these methodological flaws with the Leibovici study ad hoc. The problem with multiple endpoints is well recognized (Schulz and Grimes 2005), and there are independent reasons for being more wary about confounding in studies with small effects (Pocock, Hughes, and Lee 1987; Worrall 2010; Howick 2011). In short, while mechanistic reasoning may be useful for ruling out implausible hypotheses, it is far from clear whether it is required.

We might weaken our requirement for mechanistic reasoning still further and simply assert that spurious results (including spurious support for implausible hypotheses) are less likely when both mechanistic reasoning and evidence from comparative clinical studies support them. Such a claim might rest on the plausible premise that since each type of evidence suffers from different potential pitfalls, a hypothesis that is supported by both types of evidence is less likely to be spurious. This weaker claim may well be acceptable (although it would have to withstand the EBM objection that tightly controlled comparative clinical studies trump other evidence). However, in its altered form it can no longer be interpreted as Russo and Williamson's thesis.

To sum up, Russo and Williamson's historical and theoretical arguments are difficult to accept. In many cases, tightly controlled comparative clinical studies suffice to establish causation.

Jon Williamson (in personal correspondence) worries that I have misunderstood his position. He claims that

> the most salient mechanistic evidence is *evidence that there is an appropriate mechanism linking the putative cause to the putative effect*. This isn't evidence from a mechanism to intervention claims, but evidence that supports the claim of the existence of a mechanism. Such evidence could be got from an RCT [randomized trial], for

> example. RWT [the Russo Williamson Thesis] says that one needs evidence that there is a linking mechanism, as well as evidence that C makes a difference to E. So my main worry is that we are talking at cross-purposes!

This understanding of mechanistic evidence seems to overlap substantially with my mechanistic reasoning (a difference being that mine acknowledges that several mechanisms are often involved). Williamson's emphasis, however, is on evidence for a mechanism, whereas I emphasize the importance of the mechanism providing evidence. Unless the evidence for a mechanism is evidence "from a mechanism to intervention claims," I do not see how it helps us establish a hypothesis about the effects of an intervention. In addition, I do not understand how evidence for a mechanism can be gained from an RCT (noncircularly, i.e., without claiming that if an RCT reveals an effect that there must be a mechanism, which is a separate claim). Most importantly, when he claims that one needs a mechanism "as well as evidence that C makes a difference to E" he seems to concede that "difference-making" evidence is sufficient to establish causation.

4. Two Problems with Mechanistic Reasoning. EBM proponents are justified in being cautious about mechanistic reasoning for at least two reasons. First, biochemical mechanisms are difficult to identify. Bloodletting, placing babies to sleep on their stomachs, antiarrhythmic drugs, and many other arguably useless or harmful therapies have been adopted on the basis of reasoning from what we now believe to be wrongly identified mechanisms. Unfortunately, the more recent examples that involved more plausible mechanistic reasoning may have led to more acute harm. The apparent knowledge of what happens to some of the mechanisms under intervention lends an aura of acceptability, which, in turn, leads to more prolific use of a harmful treatment. For example, some estimate that antiarrhythmic drugs (adopted on the basis of reasoning from some but not all relevant mechanisms) killed more people every year than were killed in action during the whole of the Vietnam War (Evans, Thornton, and Chalmers 2006, 8).[1]

To illustrate why identifying relevant mechanisms is problematic, consider just one mechanism that is involved in all ingested interventions, namely, the metabolic mechanism (see fig. 3). The diagram partially represented in figure 3 is accompanied by 49 explanatory notes including the following notes: "It is still unknown, if methyl oxidation at ring B occurs before or after esterification with phytol," "In some microorganisms, cys-

1. The problem with failure to identify some relevant outcomes is often apparent when researchers use "surrogate" outcomes (such as reduction in arrhythmias) instead of patient-relevant outcomes.

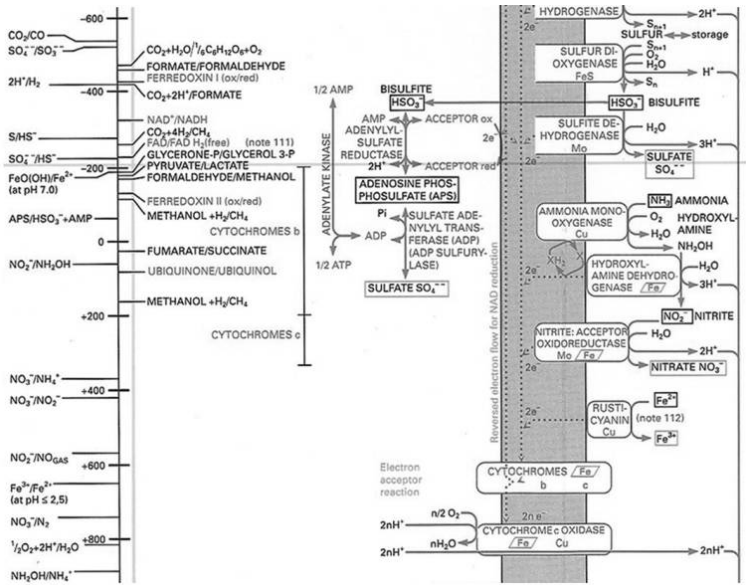

Figure 3. Small extract of the mechanism for metabolism. Source: Roche Pharmaceuticals.

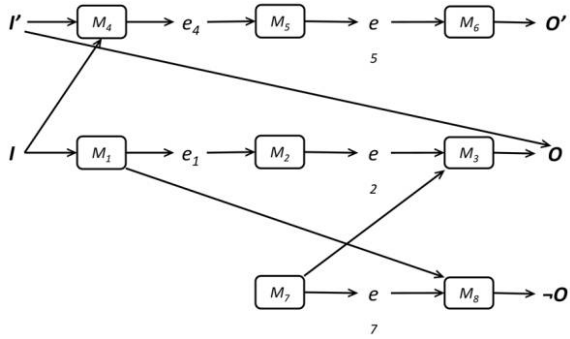

Figure 4. Paradoxical responses, where the same intervention, I, can help (O) and harm ($\neg O$) the same disorder, and outcomes are caused by more than one cause (I and I').

tathionine synthesis takes place via O-acetyl-L-homoserine," and "This reaction may also occur with the 26-hydoxylated compound" (Gerhard 1993). The partial ignorance about the metabolic mechanism means that we cannot be sure what mechanisms are eventually activated by any drug (or its metabolites) that have been swallowed. More generally, the complexity of the human body makes it difficult to identify all relevant mechanisms activated by an intervention.

The second problem with mechanistic reasoning is that it is difficult to predict how mechanisms will behave under intervention. Many philosophers of science, of course, define mechanisms as productive of regular (predictable) changes between inputs and outputs. But the regularity of mechanisms has been questioned (see DesAutels 2011). As an illustration, consider one potential source of unpredictability known as a paradoxical response, whereby the same drug has one effect in some cases and the opposite effect in others, presumably by activating different mechanisms or by causing a different response in the same mechanism (see fig. 4). Philosophers have discussed these examples for decades (Hesslow 1976). Recently, Hauben and Aronson (2006) have listed 67 drugs that sometimes worsen the condition for which they are indicated.

Less dramatic than paradoxical responses, interventions often produce unexpected harmful side effects, again, presumably by activating unsuspected mechanism(s) in some unexpected way. For instance, the first clinical trials of sildenafil found it to be ineffective for its intended use (hypertension) but quite effective at producing erections. The drug was marketed as *Viagra* and quickly became a huge commercial success. Unexpected effects of apparently "inactive" substances ("placebos") further attest to

the difficulty of predicting how bodily mechanisms behave under intervention (Golomb et al. 2010). If apparently inactive substances can have unexpected effects, then a forteriori so might "active" ones.

There is a further feature of mechanisms that is not problematic per se but adds to the abovementioned difficulties: mechanisms are almost invariably productive of stochastic relationships. The mechanism for how smoking increases the risk of lung cancer is relatively well understood, yet not all smokers contract lung cancer, nor are all lung cancers caused by smoking. Given that there are often several inferential pathways linking the intervention with the putative outcome (some with positive and others with negative effects) and that we cannot assume independence of the various mechanisms, the probabilistic nature of mechanistic relationships presents a serious problem for inferring the overall effect of an intervention fron knowledge about mechanisms.

To be sure, the problems I list above may all be epistemological, and one could object that they would be solved by deeper knowledge of mechanisms. In response, in many cases we have seen that what happens to biochemical mechanisms under medical intervention is often mistaken. This puts supporters of mechanistic reasoning in a tight spot. Either they must argue that established mechanisms in medicine are exceptional (in which case they must admit that mechanistic reasoning is unlikely to be reliable), or they must abandon the view that mechanisms provide the basis for stable and predictable causal laws. At the same time, some mechanistic reasoning is unproblematic.

5. Why EBM Proponents Should Allow a More Prominent Role for "High-Quality (Valid and Based on 'Complete' Mechanisms) Mechanistic Reasoning" in Their Evidence Hierarchies. No piece of evidence, whether from a comparative clinical study or based on underlying mechanisms, will ever be perfect. Hence, in the spirit of Fisher's hypothesis tests and Popper's falsification principle, mechanistic reasoning should be judged on the extent to which it overcomes obvious flaws outlined above. Accordingly, to be accepted, mechanistic reasoning must satisfy the following *desiderata*:

1. Knowledge of mechanisms upon which the mechanistic reasoning is based is not incomplete; that is, the mechanisms linking the intervention with the outcome have been identified and their behavior under intervention established.

A not incomplete understanding of the mechanistic chain linking the intervention with the clinically relevant outcome involves correct identification of relevant mechanisms. But identifying one inferential pathway is not enough. Mechanistic reasoning must factor in the complex

and stochastic nature of most biochemical mechanisms, so the other desiderata is:

2. The probabilistic and complex nature of the mechanisms is explicitly considered when inferring from mechanisms to any claims that a particular intervention has a patient-relevant benefit.

When the two desiderata have been met, the mechanistic reasoning in question can be judged to be of sufficiently high quality to support a hypothesis about an intervention's effects. The following real example suggests that mechanistic reasoning might sometimes suffice to support causal hypotheses without comparative clinical studies. (Somewhat ironically, this example of "high-quality" mechanistic reasoning also counts against Russo and Williamson's claim that both types of evidence are required.)

Large nodular goiters present an obstruction in the airway that impairs respiratory function. At the same time, there is strong evidence that radiotherapy shrinks goiters and that it is generally safe (Nielsen et al. 2006). Then our knowledge about the "mechanics" of breathing tells us that reducing the size of the airway obstruction will improve respiratory function. There is also strong evidence (from comparative clinical studies) that radiotherapy does not induce any paradoxical responses or severe harmful side effects. In short, there are no obvious gaps in the mechanistic knowledge linking the intervention with the patient-relevant outcome, and the possibility of serious adverse events was made unlikely by the clinical studies. Mechanistic reasoning should therefore allow us to conclude that radiotherapy will improve respiratory function, at least in the longer term.[2] However, proponents of the view that mechanistic reasoning is never valuable insisted on conducting a clinical trial of the effects of radiotherapy on goiters to improve respiratory function (Bonnema et al. 2008). They found, unsurprisingly, that radiotherapy improved respiratory function. One might even question whether the trial was ethically justified given the high-quality mechanistic reasoning.[3]

This example suggests that high-quality mechanistic reasoning can provide reliable evidence that a treatment is effective. It can be used on its own or, a forteriori, alongside evidence from comparative clinical studies to support claims about the patient-relevant benefits of medical interventions. By failing to distinguish between high- and low-quality mechanistic reasoning, EBM proponents may have overlooked an important and useful source of evidence.

2. Radiotherapy is also known to induce short-term thyroid swelling.

3. This is another example where ethics and epistemology are intertwined (Worrall 2007; Howick 2009, 2011).

6. Conclusion. The claim that Russo and Williams propound, that both mechanistic reasoning and evidence from comparative clinical studies are required, is difficult to maintain. There are many cases where patient-relevant effects of medical therapies have been established by comparative clinical studies alone (even in the face of conflicting evidence from mechanistic reasoning) and others where mechanistic reasoning without evidence from comparative clinical studies suffices. At the same time, high-quality mechanistic reasoning deserves a more prominent role in EBM hierarchies of evidence. High-quality mechanistic reasoning involves inferences from "not incomplete" mechanisms that take into account the stochastic and complex nature of mechanisms. The problem is that there is much to stand in the way of mechanistic reasoning being of high quality since there are limits to our knowledge of bodily mechanisms and their interactions.

REFERENCES

Bechtel, William, and Adele Abrahamsen. 2005. "Explanation: A Mechanist Alternative." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36:421–41.

Bernard, Claude. 1957. *An Introduction to the Study of Experimental Medicine*. New York: Dover.

Bonnema, S. J., V. E. Nielsen, H. Boel-Jorgensen, P. Grupe, P. B. Andersen, L. Bastholt, and L. Hegedus. 2008. "Recombinant Human Thyrotropin-Stimulated Radioiodine Therapy of Large Nodular Goiters Facilitates Tracheal Decompression and Improves Inspiration." *Journal of Clinical Endrocrinology and Metabolism* 93 (10): 3981–84.

Cartwright, Nancy. 2009. "Presidential Address." In *American Philosophical Association Pacific Division*. Vancouver: APA.

DesAutels, Lane. 2011. "Against Regular and Irregular Characterizations of Mechanisms." *Philosophy of Science*, in this issue.

Evans, Imogen, Hazel Thornton, and Iain Chalmers. 2006. *Testing Treatments: Better Research for Better Healthcare*. London: British Library.

Gerhard, Michal, ed. 1993. *Biochemical Pathways*. Roche Pharmaceuticals, http://www.expasy.org/cgi-bin/pathways/show_thumbnails.pl.

Gillies, D. 2005. "Hempelian and Kuhnian Approaches in the Philosophy of Medicine: The Semmelweis Case." *Studies in History and Philosophy of Biological and Biomedical Science* 36 (1): 159–81.

Glennan, Stuart S. 1996. "Mechanisms and the Nature of Causation." *Erkentnis* 44 (1): 49–71.

Golomb, B. A., L. C. Erickson, S. Koperski, D. Sack, M. Enkin, and J. Howick. 2010. "What's in Placebos: Who Knows? Analysis of Randomized, Controlled Trials." *Annals of Internal Medicine* 153 (8): 532–35.

Guyatt, G. H., A. D. Oxman, G. E. Vist, R. Kunz, Y. Falck-Ytter, P. Alonso-Coello, and H. J. Schunemann. 2008. "GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations." *British Medical Journal* 336 (7650): 924–26.

Hauben, M., and J. K. Aronson. 2006. "Paradoxical Reactions: Under-Recognized Adverse Effects of Drugs." *Drug Safety* 29 (10): 970.

Hesslow, Germund. 1976. "Two Notes on the Probabilistic Approach to Causality." *Philosophy of Science* 43:290–92.

Howick, Jeremy. 2009. "Questioning the Methodologic Superiority of 'Placebo' over 'Active' Controlled Trials." *American Journal of Bioethics* 9 (9): 34–48.

———. 2011. *The Philosophy of Evidence-Based Medicine*. Oxford: Blackwell.

Leibovici, L. 2001. "Effects of Remote, Retroactive Intercessory Prayer on Outcomes in

Patients with Bloodstream Infection: Randomized Controlled Trial." *British Medical Journal* 323 (7327): 1450–51.

Louis, Pierre Charles Alexandre. 1836. *Researches on the Effects of Blood Letting in Some Inflammatory Diseases, and on the Influence of Tartarized Antimony and Vesication in Pneumonitis*. Trans. C. G. Putnam, with preface and appendix by J. Jackson. Boston: Hillart.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67:1–25.

Mendelsohn, M. E., and R. H. Karas. 1999. "The Protective Effects of Estrogen on the Cardiovascular System." *New England Journal of Medicine* 340 (23): 1801–11.

Mill, John Stuart. 1843/1973. "A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation." In *Collected Works of John Stuart Mill*, ed. J. M. Robson, vols. 7–8. Toronto: University of Toronto Press.

Nielsen, V. E., S. J. Bonnema, H. Boel-Jorgensen, P. Grupe, and L. Hegedus. 2006. "Stimulation with 0.3-mg Recombinant Human Thyrotropin prior to Iodine 131 Therapy to Improve the Size Reduction of Benign Nontoxic Nodular Goiter: A Prospective Randomized Double-Blind Trial." *Archives of Internal Medicine* 166 (14): 1476–82.

Petitti, D. B., J. A. Perlman, and S. Sidney. 1987. "Noncontraceptive Estrogens and Mortality: Long-Term Follow-Up of Women in the Walnut Creek Study." *Obstetrics and Gynecology* 70 (3): 289–93.

Pocock, Stuart. J., M. D. Hughes, and R. J. Lee. 1987. "Statistical Problems in the Reporting of Clinical Trials: A Survey of Three Medical Journals." *New England Journal of Medicine* 317 (7): 426–32.

Rossouw, J. E., G. L. Anderson, R. L. Prentice, A. Z. LaCroix, C. Kooperberg, M. L. Stefanick, R. D. Jackson, S. A. Beresford, B. V. Howard, K. C. Johnson, J. M. Kotchen, and J. Ockene. 2002. "Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results from the Women's Health Initiative Randomized Controlled Trial." *Journal of the American Medical Association* 288 (3): 321–33.

Russo, Frederica, and Jon Williamson. 2007. "Interpreting Causality in the Health Sciences." *International Studies in the Philosophy of Science* 21 (2): 1157–70.

Schulz, K. F., and D. A. Grimes. 2005. "Multiplicity in Randomized Trials." Pt. 2, "Subgroup and Interim Analyses." *Lancet* 365 (9471): 1657–61.

Stampfer, M. J., and G. A. Colditz. 1991. "Estrogen Replacement Therapy and Coronary Heart Disease: A Quantitative Assessment of the Epidemiologic Evidence." *Preventive Medicine* 20 (1): 47–63.

Worrall, John. 2007. "Evidence in Medicine." *Compass* 2 (6): 981–1022.

———. 2010. "Evidence: Philosophy of Science Meets Medicine." *Journal of Evaluation in Clinical Practice* 116:356–62.