

Chapter 1

Introduction



Abstract This chapter introduces the idea of EBM+, which adopts the explicit requirements of EBM, to (1) make all the key evidence explicit and (2) adopt explicit methods for evaluating that evidence. EBM+ then sets out to get us better causal knowledge by explicitly integrating evidence of mechanism alongside evidence of correlation. This chapter summarises some important benefits of including evidence of mechanism, particularly given how highly idealised study populations typically are, and introduces the need to make uses of evidence of mechanism more explicit.

This book describes a number of methods that integrate the appraisal of evidence of mechanisms with other forms of evidence. While these methods are relevant to many fields where evidence is assessed (see Clarke and Russo 2016), our starting point is evidence-based medicine (EBM). The methods in this book build upon the tools already developed by EBM, by taking evidence of mechanisms into account in addition to the clinical studies that are the focus of EBM. We refer to this development as EBM+.

EBM+ Evidence of mechanisms should be integrated with evidence of correlation to better assess causal claims.

Medical practice depends fundamentally on the assessment of causal claims:

Examples of assessing causal claims in medicine.

- Identifying the causes of cancers in humans.
- Evaluating whether a medical device will lead to improved outcomes in a particular patient.
- Establishing whether a public health action will have the desired effects in the target population.
- Determining whether a medicine has a specific detrimental side-effect.
- Ascertaining the cost effectiveness of a health intervention on a target population.

Causal claims underpin evidence-based medicine, guideline development, personalised medicine, narrative medicine, and other aspects of medicine.

This book concentrates on EBM because we explicitly endorse two core EBM principles:

Two principles of EBM endorsed in this book.

1. Make all the key evidence explicit.
2. Adopt explicit methods for evaluating that evidence.

These principles have been largely responsible for the significant advances made by EBM. In particular, EBM prompted the widespread adoption of techniques for analysing data on medical interventions, with the objective of determining whether these interventions are in fact delivering the expected results.

In this book, these principles are developed with respect to evidence of mechanisms. First, evidence of mechanisms is often key evidence and needs to be made explicit. Second, evidence of mechanisms needs to be explicitly evaluated when assessing a causal claim.

1.1 What is the Key Evidence?

EBM has hitherto focused primarily on one kind of evidence for a causal claim: evidence arising from clinical studies, including randomised trials and observational studies. However, this book is motivated by the idea that evidence for causal claims in medicine cannot simply be reduced to evidence of correlation. In the philosophy of causality, the following thesis has been put forward (Russo and Williamson 2007):

Evidential pluralism. This is the thesis that one typically needs both evidence of correlation and evidence of mechanisms to establish a causal claim.

Evidential pluralism is relevant to deciding what counts as key evidence. As we shall explain, the supposition that the key evidence will be all of one type (e.g., evidence from RCTs) is not a good one. Note that this thesis about forms of evidence goes beyond the (intuitively appealing) idea that taking more evidence into account will lead to better inferences.

To develop this argument, two pieces of terminology will be helpful: [efficacy](#) and [effectiveness](#). (Technical terms are hyperlinked to their definitions. A full glossary is available in the [online appendices](#).) Although these are likely to be familiar to most readers because of their widespread use in the medical literature, our usage of these terms is broader than their usual meaning. We define these terms as follows:

Efficacy concerns the effect(s) of some intervention or exposure in a particular study population. An *efficacy claim* is a claim that the intervention or exposure has some specific effect in the study population.

Effectiveness concerns the effect(s) of an intervention or exposure in some target population of interest, such as a population of patients to be treated. An *effectiveness claim* is a claim that the intervention or exposure has some specific effect in the target population.

The term ‘efficacy’ is normally only used in the context of a *beneficial* effect of an *intervention*. However, what we have to say in this book applies equally when assessing whether an intervention causes some particular *harm*, or when assessing whether an *exposure* causes a particular harm (or, indeed, benefit). So we use ‘efficacy’ throughout this book in a more general sense, covering harms as well as benefits and exposures as well as interventions. Similarly for ‘effectiveness’.

When a relationship applies more broadly than in a study population, it is sometimes said to be *externally valid*:

External validity concerns an inference from a study to a target population. If a causal claim that holds in a study population can be extrapolated to a target population of interest, then it may be described as externally valid.

To use the terminology of Cartwright and Hardie (2012, 15), external validity concerns how we go from *knowing that something works somewhere* (efficacy) to *knowing that it will work for us* (effectiveness). Extrapolation is typically crucial for demonstrating effectiveness:

Effectiveness = efficacy + external validity. Typically, one establishes that a causal claim holds in a target population by establishing the claim in a study population and then extrapolating that claim to the target population.

The reason for proceeding to effectiveness via efficacy and external validity is that a study population is typically highly idealised, and thus differs from the target population in important ways. For example, a study population for evaluating the effectiveness of a drug might exclude those with multiple morbidities or pregnant women; a study population for evaluating the carcinogenicity of an environmental exposure might be a laboratory population of an animal model. Establishing external validity is crucial because the mechanism of action in the study population may not be particularly robust.

An **idealised population** is one which is a non-representative subpopulation of the general population. Idealised populations satisfy certain ideal experimental conditions or experience a narrowly circumscribed range of exposures.

A **robust** mechanism is one that works in the same way across a wide variety of background conditions; a **fragile** mechanism does not.

As we shall see, evidence of mechanisms is crucial to establishing both efficacy and external validity. While evidence of mechanisms is already implicit in, for example, the design of clinical trials, mechanistic studies are generally not explicitly evaluated when making policy or treatment decisions (Clarke et al. 2013, 2014). This is largely a consequence of the downplaying of mechanistic studies in the most influential EBM methods manuals (such as GRADE), owing to concerns about possible bias. While we acknowledge that there are valid concerns about biases, we regard this wholesale downplaying as a mistake. At present, evidence of mechanisms does in fact influence the evaluation of effectiveness. For example, there may be evidence that the mechanism of action in a study population is rather different from those in a target population and this difference can be taken into account when assessing the effectiveness of a drug. But this influence of evidence of mechanisms is often invisible, because it is mediated by the opinions of experts, particularly expert panel members on evidence appraisal committees. This influence is reasonable: evidence of mechanisms plays a vital role in providing evidence of effectiveness. However, the lesson of evidence-based medicine is that one needs to make evidence explicit in order to scrutinise and challenge it properly, and that one needs to make explicit the ways in which evidence is evaluated in order to improve these methods of evaluation. This book seeks to extend this evidence-based approach to include evidence of mechanisms.

Evidence of mechanisms is often produced by means other than clinical studies. In philosophy of science, much attention has been devoted to the concept of mechanism in biology and medicine, as well as in many other scientific domains (see Chap. 2 for an introduction to mechanisms). However, comparatively less attention has been devoted to the question of how evidence of mechanisms is generated and assessed, especially in the context of medical practice. This gives rise to the next major theme of this book: how should we evaluate our evidence?

1.2 The Process of Evaluating Evidence

If—in common with many of those interested in EBM—your first exposure to the methods of EBM came from the profusion of introductory articles published in the medical literature in the late 1990s (such as Sackett et al. 1996), you might get the impression that the quality of a piece of clinical research could be determined

with relatively straightforward judgements of the methodology used in the research. Was the research randomised? Did the authors use intention-to-treat analysis? Had the statistical analysis produced a significant result? Unless these conditions were jointly satisfied, the research was of very low quality. And if they were satisfied, then it was likely that the work was of high quality, and should be used as a guide to practice—unless very serious provisos were detected (such as research misconduct).

This is because the evaluation of evidence in early EBM was about describing the methods used to produce that evidence. This placed the onus of judging the quality of a piece of research largely on the reader. In turn, this led to an emphasis on critiquing research methods as a proxy for judging the quality of research (Greenhalgh 2014, 28). Concerns about bias were given priority, and this heightened scrutiny of research methods has been the major defence against biased research.

However, critiquing research methods (rather than the details of a specified piece of research) is only possible because—for all the many complications of doing clinical research—many individual clinical studies share the same fundamental design. This means that shared ways of evaluating quality can be fairly easily learned and applied by health scientists, with the reasonable expectation that these simple methods are effective in stripping out biased research.

There is a fallacy here. Evaluating a small number of indicators did much of the work in downgrading biased research, and it did it in an efficient and simple way. Yet that is not to say that these techniques worked without any judgement on the part of the evaluator. Nor did these techniques work flawlessly. Although some research designs are more prone to bias than others, it does not automatically follow that, for instance, *all* non-randomised research is intrinsically biased. To use the terminology devised by Kahneman (2011), this kind of evaluation is a kind of system 1 thinking: fast and easy, but prone to faults. We are rightly suspicious of other kinds of system 1 thinking because of its propensity to bias. But sometimes speed is preferable to accuracy, and system 1 thinking may often be good enough. And we might choose to evaluate, for instance, clinical studies in this system 1 manner because there is a common structure of clinical research that allows us to make good enough judgements about their likely quality (Kahneman 2011). If EBM was a useful first approximation to evidence evaluation, then EBM+ is intended as a second, improved, approximation.

The same assumptions about commonality of methods do not seem to apply to evidence of mechanisms. It is hard to think of a field with more methodological diversity than contemporary bioscience research. For example, computer simulations, 31-P NMR, mass spectrometry, knockout studies and immunofluorescence do not exhaust the space of research strategies that have been used to understand a single protein (Mitchell and Dietrich 2006). And so we do not offer, in this book, a tool capable of evaluating all of this research in a substance-blind manner. We note in passing too that the presence of candidate indicators for clinical studies (such as intention to treat analysis, randomisation, or trial registration) that have been touted as ensuring that a piece of research can be accepted without question do no such thing, although they are individually helpful to an expert judge of clinical evidence. We need to judge evidence, and the methods and tools provided here are an aid to judgement, rather than a replacement for it.

1.3 Our Approach to Evaluating Evidence

The approach to evaluating evidence that is developed in this book can be traced back to work of Russo and Williamson (2007), who put forward an account of evidential pluralism in medicine. Williamson (2018) offers a recent defence of evidential pluralism.

Evidential pluralism in medicine is not a new idea. For instance, the causal indicators put forward by Hill (1965) can be viewed as a version of evidential pluralism. Several of Hill's indicators of causality are good indicators of mechanisms, while others are good indicators of correlation. We discuss Hill's indicators, and explain how our approach improves over them, in Chap. 6; see also Williamson (2018b).

The methods for evidence evaluation that we set out in later parts of the book all require judgement on the part of the user (Kelly and Moore 2012; Montgomery 2005). We do not pretend that there is a shibboleth or an algorithm that determines the excellence (or otherwise) of a piece of evidence of mechanisms. All evaluations of quality of evidence are fallible. With this work, we hope to reach those readers interested in combining practical methods for evidence evaluation with philosophical analysis.

References

- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press.
- Clarke, B., & Russo, F. (2016). Causation in medicine. In J. Marcum (Eds.), *Companion to contemporary philosophy of medicine* (pp. 297–322). Bloomsbury.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). The evidence that evidence-based medicine omits. *Preventive Medicine*, 57(6), 745–747.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, 33(2), 339–360.
- Greenhalgh, T. (2014). *How to read a paper: The basics of evidence-based medicine*. New York: Wiley.
- Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58, 295–300.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin.
- Kelly, M. P., & Moore, T. A. (2012). The judgement process in evidence-based medicine and health technology assessment. *Social Theory and Health*, 10(1), 1–19.
- Mitchell, S. D., & Dietrich, M. R. (2006). Integration without unification: An argument for pluralism in the biological sciences. *The American Naturalist*, 168(S6), S73–S79.
- Montgomery, K. (2005). *How doctors think: Clinical judgment and the practice of medicine*. Oxford University Press.
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21(2), 157–170.

- Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal*, 312, 71–2.
- Williamson, J. (2018a). Establishing causal claims in medicine. *International Studies in the Philosophy of Science*, in press. Preprint available at <http://blogs.kent.ac.uk/jonw/>.
- Williamson, J. (2018b). Establishing the teratogenicity of Zika and evaluating causal criteria. *Synthese*, in press. Preprint available at <http://blogs.kent.ac.uk/jonw/>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

