

Chapter 3

How to Consider Evidence of Mechanisms: An Overview



Abstract This chapter introduces how to assess evidence of mechanisms, explaining a summary protocol for use of evidence of mechanisms in assessing efficacy, then external validity (developed theoretically in Part III, with tools for implementation offered in Part II). An outline of quality assessment—of a whole body of evidence, rather than individual studies—is given. The chapter finishes with a brief introduction to the ideas developed in Part III: gathering evidence of mechanisms (Chap. 5); evaluating evidence of mechanisms (Chap. 6); and using evidence of mechanisms to evaluate causal claims (Chap. 7).

This section summarises the overall approach taken in this book. It develops some of the more practical issues raised in the introduction (Chap. 1) and begins to attach these to the more theoretical discussions found in Part III. We start with an overview of the way in which effectiveness can be evaluated. As discussed above, effectiveness can be evaluated by evaluating efficacy and external validity. A translation of the core ideas of this chapter to other arenas of practice, such as social policy, is readily possible—although we do not attempt this here in the interest of clarity.

3.1 Questions to Address

The following protocol can be used to test a causal claim:

Efficacy

Does the effect size and quality of clinical studies establish that the observed correlation is causal?

Yes? Efficacy is established.

No?

- Evaluate other evidence for the claim that there exists an appropriate mechanism that can explain the observed correlation.
 - What are the hypothesised mechanisms?
 - How well confirmed is each such mechanism? What are the gaps? How well confirmed is each feature (process, entity, activity and organisational feature) of the mechanism?
 - Can the mechanism account for the full effect size? Are there counteracting mechanisms? What is the evidence that the influence of any counteracting mechanisms is less than that of the proposed mechanism?
- Evaluate other evidence to rule in or out other explanations of the correlation. Are any remaining explanations better confirmed than the hypothesis that the correlation is causal?

Efficacy is established if one can establish, in the study population, the existence of a correlation and the existence of a mechanism that can explain this correlation.

External validity

Do clinical studies directly establish a suitable association and mechanism in the target population?

Yes? Effectiveness is established.

No?

- Evaluate the claim that the mechanism of action is sufficiently similar in the target and study populations.
- Evaluate the claim that in the target population, any counteracting mechanisms that are not also present in the study population do not mask the effect of the mechanism of action.
- Evaluate other evidence for a correlation in the target population.

External validity is established if one can establish similarity of relevant mechanisms in the study and target populations, and thereby establish, in the target population, the existence of a correlation and the existence of a mechanism that can explain this correlation.

In the case of efficacy, it is rare that clinical studies *alone* establish that the observed correlation is causal in the study population. Clinical research does not (generally) take place in isolation from basic science research. Many aspects of the design and interpretation of clinical trials—such as the choice of outcome measures, therapeutic regimes compared, and patient recruitment criteria—are influenced by evidence of mechanisms. Thus, even in the absence of complete knowledge of the underlying mechanisms, evidence of mechanisms contributes to establishing efficacy (Illari 2018). This is also true with respect to external validity, where it is almost never the case that clinical studies in the study population will directly establish both a suitable association and mechanism that will apply to the target population. Rather, external validity inferences proceed in one of the following ways (Parkkinen and Williamson 2017):

1. By identifying and comparing key details of the mechanisms in the study and target populations.
2. Inductively, by observing a similar effect in many different experimental populations and generalizing from these to the target population (Wilde and Parkkinen 2017).
3. Phylogenetically, by identifying the mechanism in the study population, and then inferring that the mechanisms in the study and target population are similar due to shared ancestry of the populations. The greater the degree of isolation between the target population and the study population, the less reliable this inference will be.
4. By means of a robustness argument: showing that the mechanism in the study population is so robust, and differences between the study and target populations are so minor, that the mechanism of action will also obtain in the target population.

Thus, for both efficacy and external validity one typically needs to consider evidence of mechanisms arising from sources other than the clinical studies that establish a correlation in the study population. This means that those who evaluate evidence will generally need to consider [mechanistic studies](#), in addition to clinical studies, in order to make causal judgements.

Of course, some features of a putative mechanism may already be well established, in which case there will usually be no need to revisit the evidence for those features. Other features will be more contentious. It is only by explicitly identifying these features and the evidence that pertains to them that one can critically appraise a proposed mechanism.

3.2 Quality of Evidence and Status of Claim

Quality of evidence. Evidence for various claims can be ranked by quality. We distinguish three main kinds of claim: claims about correlation, claims about mechanisms and causal claims (including claims about efficacy and claims about external validity). We use the scale in Table 3.1 to rank the quality of this evidence.

Table 3.1 Quality levels of evidence, based on Atkins et al. (2004)

Quality level	Interpretation
High	Further research is highly unlikely to have a significant impact on our confidence in the claim
Moderate	Further research is moderately unlikely to have a significant impact on our confidence in the claim
Low	Further research is moderately likely to have a significant impact on our confidence in the claim
Very low	Further research is highly likely to have a significant impact on our confidence in the claim

Note that this ranking system evaluates the *total body of evidence* pertaining to the claim in question. This is in sharp contrast to other EBM methods that evaluate single studies in isolation.

This approach to ranking quality on the basis of stability of confidence can be found in the original GRADE framework (see Guyatt et al. 2008). According to this sort of approach, establishing a causal claim requires confidence in the *stability* of that causal claim, in addition to confidence in the nature of the claim itself. We should emphasise that the interpretation of each category concerns the *in principle possibility* of obtaining further research that changes confidence in the claim. A brief example will be helpful here. Suppose current evidence warrants 75% confidence in a causal claim. One then learns that there is further evidence which warrants a 25% change in confidence, but one does not know the direction of this change. i.e., one does not know whether this new evidence warrants 50% confidence or 100% confidence. The 75% confidence is not sufficiently stable for the claim to be considered established or even provisionally established. This is because future evidence may be likely to decide between the 50 and 100% confidence, leading to a large change in confidence either way.

GRADE later changed their interpretation of quality levels, dropping reference to the likelihood that further evidence will change confidence in the claim (Balshem et al. 2011, Table 2; Hulcrantz et al. 2017). This was because of concerns about the situation in which further evidence is unlikely to be obtained in practice: if further research is unlikely to be carried out then further research is unlikely to have an impact on our confidence in the causal claim in question. This change is unnecessary: as noted above, the key question is whether evidence can *in principle* be obtained to significantly alter confidence in the claim. In short, just because ethical or practical considerations make it very unlikely that further research on a particular claim will be carried out, that does not imply that current evidence is high quality.

Table 3.2 Status of a claim

Status	Interpretation
Established	A claim is established when community standards are met for adding the claim to the body of evidence—i.e., for granting the claim and treating it as evidence for other claims In order to establish a claim, evidence must warrant a high level of confidence in the claim and this evidence must itself be high quality
Provisionally established/ provisional	Moderate quality evidence warrants a high level of confidence in the claim
Arguably true/ arguable	The claim is neither established nor provisionally established, but evidence of at least moderate quality warrants significantly more confidence in the claim than in its negation, or low quality evidence warrants a high level of confidence in the claim
Speculative	A claim is speculative if it falls into none of the other categories
Arguably false	The claim is neither ruled out nor provisionally ruled out, but evidence of at least moderate quality warrants significantly more confidence in the negation of the claim than in claim itself, or low quality evidence warrants a high level of confidence in the negation of the claim
Provisionally ruled out	Moderate quality evidence warrants a high level of confidence in the negation of the claim
Ruled out	A claim is ruled out when community standards are met for adding the negation of the claim to the body of evidence In order to rule out a claim, high quality evidence must warrant a high level of confidence in the negation of the claim

Status of claim. In addition to the quality of the evidence, we shall also be concerned with the status that the evidence confers on the claim under consideration. The status of a claim will be measured on the scale depicted in Table 3.2.

Note that this table invokes two separate levels: the quality level applies to the total evidence, while the level of confidence applies to the claim in question. The status of the claim depends on both the quality of the evidence as well as the degree of confidence that the evidence warrants.

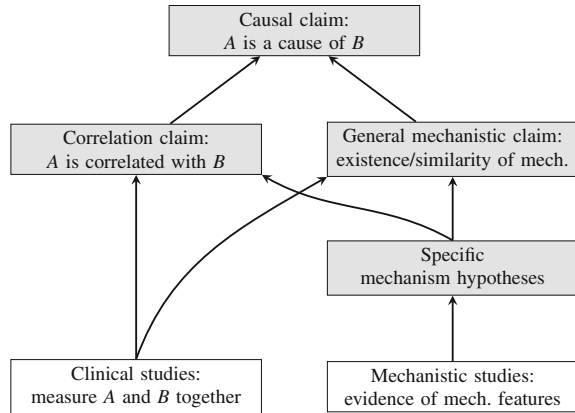
We will see shortly that the status of a causal claim will depend on the status of a correlation claim (assessed, e.g., by using the GRADE system) together with the status of a mechanism claim (assessed by the procedures outlined in Chap. 6).

Appendix B provides a simple probabilistic interpretation of the notion of quality and status developed in this section.

3.3 Overall Approach

Figure 3.1 depicts the evidential relationships linking the concepts of this book; cf. Williamson (2018b). A claim that *A* is a cause of *B* is assessed by evaluating

Fig. 3.1 The evidential relationships employed in this book. See Williamson (2018b)



two further claims. The first—the **correlation** claim—is the claim that A and B are appropriately correlated. The second is the **general mechanistic claim**. In the case of efficacy, this is the claim that there exists an appropriate mechanism linking A and B that can explain B in terms of A and that can account for the extent of the correlation. There are two ways of confirming this general mechanistic claim: either via clinical studies which find a correlation that can only be explained by the general mechanistic claim being true, or by identifying key features of the actual mechanism of action, which are confirmed by mechanistic studies. In the case of external validity, the general mechanistic claim is the claim that the mechanisms of action in the study and target population are sufficiently similar. Again, this can be confirmed either by clinical studies on both populations that find similar correlations, or by ascertaining key features of the mechanism of action in each population and finding that these are similar. In addition, clinical studies provide good evidence of correlation, and, in certain circumstances, an established mechanism of action can also provide good evidence of correlation (Williamson 2018a, Sect. 2.2).

There is a **correlation** between two variables A and B if these two variables are probabilistically dependent, i.e., $P(B|A) \neq P(B)$. In many situations where a causal relationship is being assessed, the correlation claim of interest is the probabilistic dependence of A and B conditional on some set of a priori potential confounding variables. A confounding variable is a variable correlated with both A and B , such as a common cause of A and B . Note that ‘correlation’ is sometimes used to refer to a linear dependence; here we use the term in the more general sense to refer to any probabilistic dependence.

Specific mechanism hypothesis. This is a hypothesis of the form: a specific mechanism with features F links the putative cause to the putative effect.

In contrast, other current EBM methods for evidence appraisal focus almost exclusively on the evaluation of clinical studies, i.e., on the two arrows at the bot-

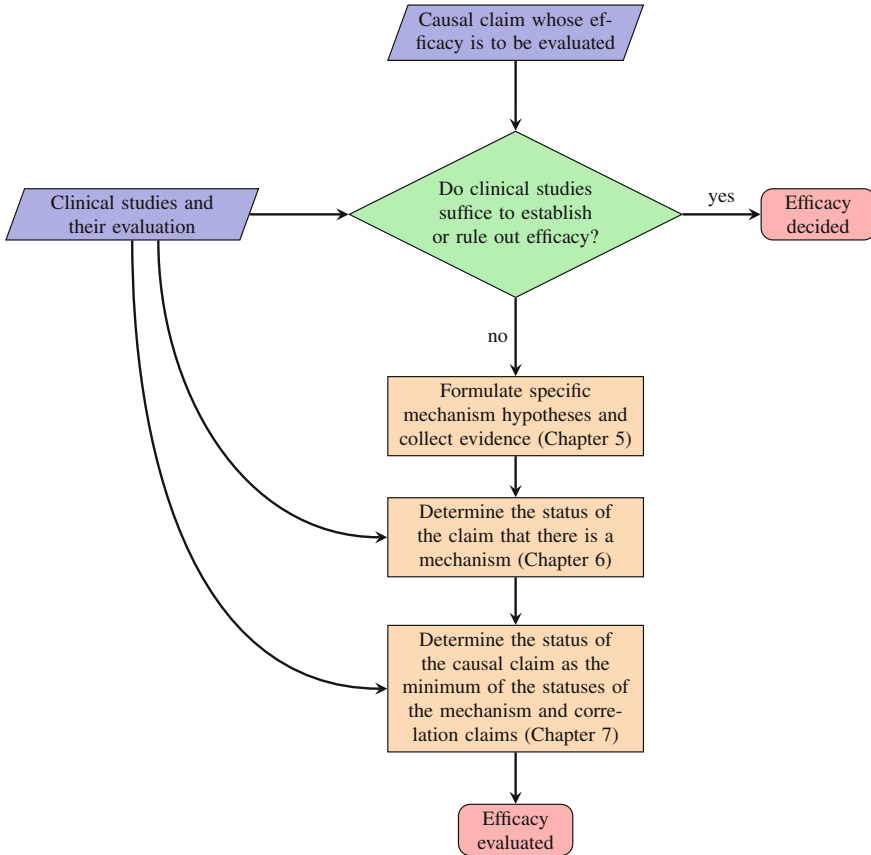


Fig. 3.2 Evaluating efficacy

tom left of Fig. 3.1. Moreover, they tend to conflate these two arrows—they do not distinguish the role of clinical studies in evaluating a correlation claim from their role in determining whether there is some underlying mechanism of action. Once these two roles are separated, it is clear that mechanistic studies also need to be appraised when evaluating the latter general mechanistic claim. This is the **evidential pluralism** introduced in Sect. 1.1.

Two flowcharts summarise the overall approach. Figure 3.2 depicts the workflow when evaluating efficacy. The second flowchart, Fig. 3.3, applies to the evaluation of external validity. In each case there are three principal steps: gathering evidence of mechanisms; evaluating evidence of mechanisms; and using evidence of mechanisms to evaluate causal claims. Procedures for implementing the three steps are developed in Chaps. 5, 6 and 7 respectively. The main ideas can be summarised as follows.

Gathering evidence of mechanisms (Chap. 5). It is typically more difficult to find evidence of mechanisms in the literature than it is to find relevant evidence of cor-

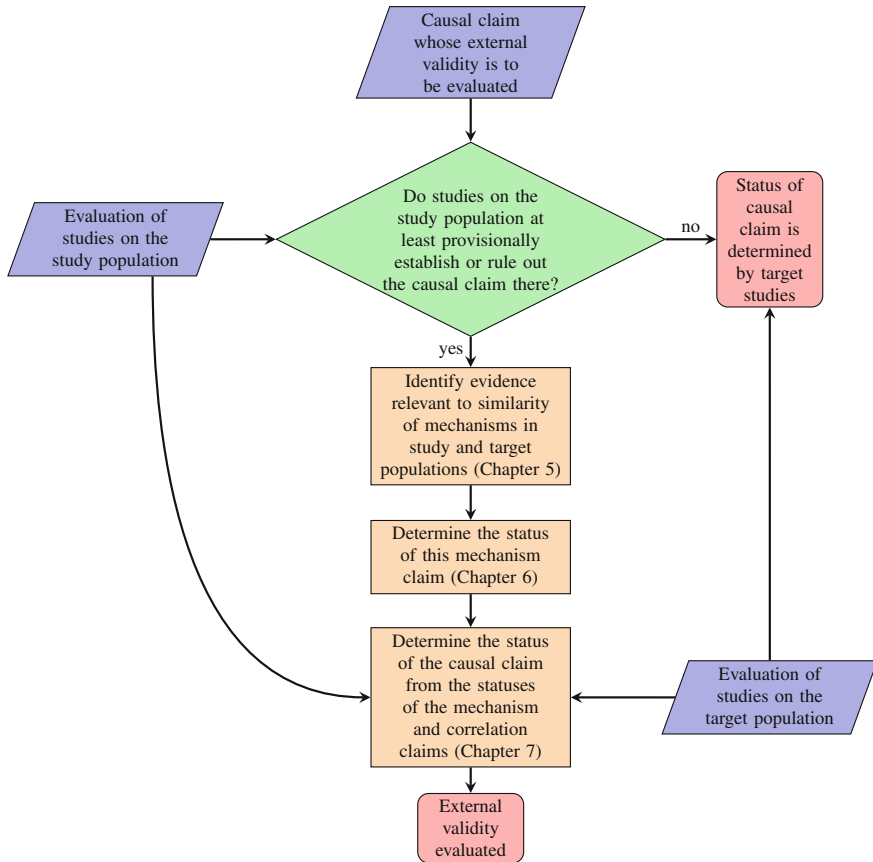


Fig. 3.3 Evaluating external validity

relation. This is because evidence of mechanisms is characteristically produced by mechanistic studies, and there are a large number of diverse types of mechanistic study (Smith et al. 2016). This makes the process of recognising good evidence more difficult, because an investigator is likely to be unfamiliar with the details of all the possible kinds of research that might be relevant to a clinical outcome. Historically, as Evans (2002) has argued, database indexing practices for these studies have tended to be unsystematic in comparison with those for clinical studies. Arguably, this has contributed to a tendency to overlook or entirely ignore evidence of mechanisms that arises from sources other than clinical studies.

However, as explained above, such evidence of mechanisms is often crucial to establishing efficacy and external validity. Given this, the difficulties in gathering evidence of mechanisms need to be overcome. As a first step towards overcoming the difficulties, we propose a five-step strategy for identifying evidence of mechanisms, a strategy that in part relies upon existing evidence of mechanisms:

1. *Identify*: Identify a number of specific mechanism hypotheses.
2. *Formulate*: For each specific mechanism hypothesis, formulate a number of review questions.
3. *Search*: Use these review questions to search the literature.
4. *Refine*: Identify the evidence most relevant to the mechanism hypothesis in question by refining the results of this search.
5. *Present*: Present the evidence relevant to the mechanism hypothesis.

This strategy is intended to help overcome some of the practical difficulties of identifying evidence of mechanisms—difficulties which may prevent practitioners from considering all the evidence. We develop this strategy in more detail in Chap. 5. We have also provided a series of tools in Part II that help users conduct certain parts of this process in specified areas of practice.

Evaluating evidence of mechanisms (Chap. 6). In evaluating the quality of mechanistic evidence, the following questions are likely to be most helpful.

1. *How well established and understood are the methods by which the evidence (of existence of a mechanism or some of its features) was produced?*
Well established methods whose functioning and potential biases are properly understood and which can be calibrated against other well established methods typically provide higher quality evidence than methods that rely on novel techniques that cannot be calibrated against better understood methods.
2. *Can the item of evidence be produced by independent methods?*
Employing several detection techniques and checking their results against each other is a common way to distinguish experimental artefacts from valid results. (The greater the number of independent methods that can confirm a result, the higher the quality of an item of evidence.)
3. *Are the model systems that are used in experimental research well characterised?*
Model systems do not usually exactly reproduce the relevant human mechanisms. Have the relevant differences been characterised for the system(s) used in this research?
4. *Can the mechanism be observed operating in many different background contexts?*
The more robust a mechanism is against variation in background conditions, the less likely it is that inferences based on evidence of the mechanism will err because of unknown contextual factors interfering with the mechanism. Demonstrable robustness of the mechanism itself thus makes for higher quality evidence.

Sections 6.1 and 6.2 describe a procedure for evaluating the quality of mechanistic studies that is broken down to three steps:

1. Evaluating methods
2. Evaluating the implementation of methods
3. Evaluating results

The **status** of the general mechanistic claim is then assigned as follows. A mechanism to account for efficacy can be considered *established* in two ways. First, when high quality clinical studies exhibit a substantial correlation that is not explainable by, e.g., confounding or bias. Alternatively, when there are high quality mechanistic studies that confirm all the crucial component features of the mechanism. A hypothesised mechanism for efficacy is considered *ruled out* when there is high quality evidence against the existence of the component features of the mechanism. A mechanism may also be ruled out if high quality clinical studies consistently fail to show results one would expect if the mechanism was operating as hypothesised. A mechanism to account for external validity is considered *established* when high quality evidence establishes the similarity of all the crucial components of the mechanism in the study and target populations. A mechanism hypothesised to account for external validity is considered *ruled out* when there is high quality evidence of dissimilarity of mechanisms between the study and target populations. The more gaps or inconsistencies there are in the evidence base for a particular claim about a mechanism, the lower its status.

There are other useful status indicators that require slightly more careful judgement. *Provisionally established* claims admit some gaps in the evidence base, but require overall a good amount of high quality evidence. *Arguable* claims have evidence in their support that is either moderate quality or that has important gaps. *Speculative* claims are supported by evidence that shows mixed results, or have little evidence in their support beyond theoretical intuition or speculation.

These issues are explained in more detail in Chap. 6.

Using evidence of mechanisms to evaluate causal claims (Chap. 7). Having ascertained the status of a correlation claim and relevant mechanism claims, one can use these to determine the status of the causal claim of interest. This process, which is explored in Chap. 7, may be summarised as follows.

In order to establish efficacy, one needs to establish that the putative cause and effect are correlated and that there is a mechanism that can account for this correlation. More generally, one can take the status of a causal claim to be the minimum of the status of the correlation claim and the status of the general mechanistic claim. For instance, if a correlation is arguable but the existence of any underlying mechanism is provisionally ruled out, then the causal claim itself is provisionally ruled out.

Turning to external validity, the situation is more complicated because one needs to consider (i) evidence for the causal claim obtained directly on the target population, (ii) evidence for efficacy in the study population, and (iii) evidence of similarity of mechanisms between the study and the target populations. Evidence directly about the target may be boosted (or undermined) by observing that efficacy does (or does not) hold in a study population that shares similar mechanisms with the target population. Table 7.1 combines the status of the causal claim in the target with the status of efficacy in the study and the status of the claim that the mechanisms in the target and the study are similar.

References

- Atkins, D., Best, D., Briss, P. A., Eccles, M., Falck-Ytter, Y., Flottorp, S., et al. (2004). Grading quality of evidence and strength of recommendations. *BMJ*, 328(7454), 1490–1490.
- Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., et al. (2011). Grade guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, 64(4), 401–406.
- Evans, D. (2002). Database searches for qualitative research. *Journal of the Medical Library Association*, 90, 290–293.
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., et al. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, 336(7650), 924–926.
- Hultcrantz, M., Rind, D., Akl, E. A., Treweek, S., Mustafa, R. A., Iorio, A., et al. (2017). The grade working group clarifies the construct of certainty of evidence. *Journal of Clinical Epidemiology*, 87(Supplement C):4–13.
- Illari, P. M. (2018). Who's afraid of mechanisms? *British Journal for Philosophy of Science*.
- Parkkinen, V.-P., & Williamson, J. (2017). Extrapolating from model organisms in pharmacology. In A. La Caze & B. Osimani (Eds.), *Uncertainty in pharmacology: Epistemology, methods, and decisions*. Dordrecht: Springer.
- Smith, M. T., Guyton, K. Z., Gibbons, C. F., Fritz, J. M., Portier, C. J., Rusyn, I., et al. (2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. *Environmental Health Perspectives*, 124, 713–721.
- Wilde, M., & Parkkinen, V.-P. (2017). Extrapolation and the Russo–Williamson thesis. *Synthese*. <https://doi.org/10.1007/s11229-017-1573-y>.
- 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.

