

Chapter 7

Using Evidence of Mechanisms to Evaluate Efficacy and External Validity



Abstract Previous chapters in Part III develop accounts of how to gather and evaluate evidence of claims about mechanisms. This chapter explains how this evaluation can be combined with an evaluation of evidence for relevant correlations in order to produce an overall evaluation of a causal claim. The procedure is broken down to address efficacy, external validity, and then the overall presentation of the claim.

In this chapter, we move from claims about mechanisms to causal claims, i.e., claims of efficacy and external validity. As we have seen in Chap. 6, in order to establish efficacy, one needs to establish both the claim that there is a correlation between putative effect and putative cause and the claim that there is a mechanism connecting the putative effect and cause that can account for the size of the observed correlation. Sect. 7.1 explains how these two types of evidence can be combined to evaluate the status of an efficacy claim. For purposes of clinical or public health decision making one often wants to make inferences about effectiveness, i.e., about causality in target populations other than the study population. Besides evidence directly about the target population, evidence of mechanistic similarity between the target populations and study populations for which efficacy has already been evaluated may be relevant to the status of the causal claim in the target population. We deal with this question of external validity in Sect. 7.2.

7.1 Efficacy

Here we address the question of how to combine evaluations of a general mechanistic claim and a correlation claim in order to evaluate a claim of effectiveness.

General mechanistic claim. We have seen (in Chap. 6) that the status of the claim that there is a mechanism connecting putative cause and effect is assessed along two different dimensions:

1. Is clinical study evidence strong enough to make it plausible that there is a mechanism that can account for the size of the correlation?
2. Is there a specific mechanism hypothesis and is the existence of the crucial features of *that* mechanism hypothesis established?

Correlation claim. The correlation claim is the claim that there is a [correlation](#) between the putative cause and effect, conditional on plausible confounders. Note that mechanistic evidence and results from previous clinical studies may rule in some variables as plausible confounders. Mechanistic evidence may also speak to the question whether a certain clinical study is well-conducted and properly controlled for these confounding variables. Given that one has settled on both a set of potential confounders and an assessment of the quality of the design of the relevant studies, deciding whether the putative cause and effect are correlated is a purely statistical question. A meta-analysis, for instance, of relevant studies yields an estimate for the size of the correlation and corresponding confidence interval and p -value. The status of the correlation claim then depends on the width of the confidence interval, the size of the p -value, and the heterogeneity of the studies evaluated. A low p -value may, for instance, lead to a high status of the correlation claim.

Efficacy claim. To obtain the status of an efficacy claim, we combine the status of the corresponding general mechanistic claim with the status of the corresponding correlation claim. Efficacy is established just when it is established that there is a correlation and that there is some mechanism which can account for this correlation (Russo and Williamson 2007; Illari 2011; Clarke et al. 2013, 2014). More generally, the status of the causal claim can be taken to be the minimum of two statuses: the status of the correlation claim and the status of the general mechanistic claim:

Status of an efficacy claim. The status of the claim A is a cause of B is the minimum of:

1. the status of the claim that A and B are appropriately correlated, and
2. the status of the claim that there is an appropriate mechanism linking A and B that can account for this correlation.

Hence, a causal claim cannot have a higher status than both the correlation claim and the general mechanistic claim (see discussions in (Russo and Williamson 2007, 2011, 2012; Russo 2011; Clarke and Russo 2016, 2017)). To give an example, efficacy is provisionally established if the existence of a correlation is established or provisionally established and the existence of a mechanism that can account for the correlation is provisionally established. Equally, efficacy is provisionally ruled out if a correlation is provisionally ruled out and if the existence of a mechanism that can account for the correlation is provisionally ruled out or of higher status.

Before turning to external validity, we discuss a potential source of confusion:

Digression: reinforced concrete. In the framework set out above, there are two separate distinctions in play. First, there is the distinction between evidence of correlation and evidence of mechanisms (Illari 2011). This distinction is core to the approach

taken in this handbook: the claim that *A* is a cause of *B* is evaluated according to how strongly evidence of correlation supports the claim that *A* and *B* are appropriately correlated, and how strongly evidence of mechanisms supports the general mechanistic claim that there is a mechanism linking *A* to *B* that can account for the correlation. Second, there is a distinction between **clinical studies** (which repeatedly measure *A* and *B* together) and **mechanistic studies** (which investigate the details of a putative mechanism linking *A* and *B*). It is important to note that these two distinctions do not align. Both clinical and mechanistic studies can provide evidence of correlation (though clinical studies often provide better evidence of correlation than mechanistic studies). Similarly, both clinical and mechanistic studies can provide evidence of mechanisms (although mechanistic studies often provide better evidence). See Fig. 3.1. Moreover, there are situations in which a causal claim can be established on the basis of clinical studies alone, as explained in Sect. 2.3 and Chap. 6.

Clinical studies and mechanistic studies can be mutually reinforcing. Consider an analogy to reinforced concrete, which is formed by placing steel grids into concrete (Clarke et al. 2014). Concrete has high resistance to compressive stresses but fractures under tension. Steel, however, has high strength in tension. So, if steel is placed in concrete to produce reinforced concrete, we get a composite material where the concrete resists the compression and the steel resists the tension. The combination of two different materials produces a material that is much stronger than either of its components. In the same way, combining clinical studies with mechanistic studies produces much stronger overall evidence of efficacy than would either type of evidence on its own, because they compensate for each other's weaknesses. For instance, clinical studies can rule out masking: masking occurs when one or more counteracting mechanisms cancel out the effect of the mechanism of action. On the other hand, mechanistic studies can rule out confounding.

The following scenarios illustrate the idea of reinforced concrete.

Scenario 1. Suppose, for instance, that many well conducted RCTs consistently show a correlation between the putative cause and effect and that bench research provides only very low quality evidence for the general mechanistic claim that there exists a mechanism that can account for the size of the correlation. In this case, it might seem that the correlation is established and the existence of the mechanism is speculative. In which case, efficacy is only speculative. However, this misrepresents the evidence for the general mechanistic claim. It confuses evidence obtained *only* by bench research with total evidence of mechanisms from all sources. Recall from Sect. 6.3 that clinical studies may also yield evidence relevant to the general mechanistic claim that there exists a mechanism—see Joffe (2011) and Williamson (2018, Sect. 2.1). In the above example, the RCTs, when combined with the bench research, can yield a status for the general mechanistic claim that is higher than speculative—an application of the reinforced concrete metaphor. Accordingly, the efficacy claim will have a status higher than speculative.

Scenario 2. Suppose low quality clinical studies suggest that there is a correlation. Suppose too that high quality mechanistic studies support key aspects of a specific mechanism hypothesis, but that the possibility of a counteracting mechanism cannot be ruled out. In this case, it is not clear that the proposed mechanism of action can

account for the observed correlation, and the general mechanistic claim will not be established. Subsequently, high quality clinical studies are carried out and determine that the net correlation is indeed positive. These studies provide evidence that any counteracting mechanism fails to totally mask the effect of the mechanism of action. The total body of evidence may now suffice to establish the general mechanistic claim (see Sect. 6.3). In this scenario, clinical studies reinforce mechanistic studies when evaluating the general mechanistic claim.

Scenario 3. Suppose certain clinical studies provide low quality evidence of a correlation. One might think that the key concern is confounding, so that when there is high quality evidence of mechanisms that rules out confounding, efficacy is established. However, confounding is not the only problem that arises with low quality evidence of correlation. There is also the problem that the observed correlation may not correspond to a correlation in the underlying data-generating probability distribution. In order to establish efficacy, one needs to establish that there is a genuine correlation in the underlying distribution. Hence, without high quality evidence of correlation, efficacy cannot be established.

Scenario 4. Suppose that initially, certain clinical studies provide low quality evidence of a correlation. Suppose that in this case, it is clear that the studies identify a genuine correlation conditional on certain potential confounders, but that not all plausible confounders have been controlled for. The key concern here, then, is confounding. For instance, there might be a large number of epidemiological studies all showing a correlation between putative cause and effect, but where each study fails to control for some particular variable which may be a confounding variable. Now, if there is also high quality evidence of mechanisms that rules out this variable as a confounder, efficacy is established. In this case, the mechanistic studies boost the status of the correlation claim, to established. In this case, then, the overall status is established.

7.2 External Validity

When mechanisms within a study population and the target population are sufficiently similar, one can extrapolate an efficacy claim from the study population to the target population. In this section, we show how to combine evidence of efficacy obtained directly on the target population with evidence obtained by extrapolation from a study population.

Three assessments feed into the evaluation of effectiveness:

1. *Efficacy in the target population.* Although studies performed directly on the target population will normally be less conclusive than those performed on the study population, they can form the basis of a preliminary evaluation of efficacy in the target population. The preliminary status of the causal claim can be determined as set out in Sect. 7.1.

Table 7.1 Determining the status of the causal claim in the target population given the status of the causal claim in the study population, the status of the claim that the mechanisms of action in study and target are similar, and the status of the causal claim in the target population on the basis only of studies carried out on the target population

| | | Causation in study population + similarity of mechanism in target and study | | | | |
|--|---------------------------|---|---|---------------------------|---|---------------------------|
| | | Established + established | Provisionally established + established; or Established + provisionally established | Other combinations | Provisionally ruled out + established; or Ruled out + provisionally established | Ruled out + established |
| Causation in target from target studies | Established | Established | Established | Established | Established | Provisionally established |
| | Provisionally established | Established | Provisionally established | Provisionally established | Provisionally established | Arguable |
| | Arguable | Established | Provisionally established | Arguable | Speculative | Speculative |
| | Speculative | Established | Arguable | Speculative | Arguably false | Ruled out |
| | Arguably false | Speculative | Speculative | Arguably false | Provisionally ruled out | Ruled out |
| | Provisionally ruled out | Arguably false | Provisionally ruled out | Provisionally ruled out | Provisionally ruled out | Ruled out |
| | Ruled out | Provisionally ruled out | Ruled out | Ruled out | Ruled out | Ruled out |

2. *Efficacy in the study population.* The status of efficacy in the study population can also be determined by considering the procedure of Sect. 7.1.
3. *Similarity of mechanisms in the study and target populations.* The status of the general mechanistic claim relevant to external validity (i.e., the claim that the mechanisms of action are sufficiently similar in study and target) can be determined as indicated in Sect. 6.3.

To obtain a final status for efficacy in the target, one can combine the preliminary status in the target population with the status of efficacy in a study population, provided that study and target population share similar mechanisms of action. The status of the causal claim about the target population may be increased (respectively, decreased) by observing that efficacy does (respectively, does not) hold in a study population that is similar to the target population. In this case, causal claims are extrapolated from the study population to the target population.

Table 7.1 shows how the **status** of the causal claim in the target population can be determined from the above three assessments. To change the preliminary status of an efficacy claim given by studies directly on the target population, all evidence of causation in the study population and of similarity of mechanisms needs to be of at least moderate quality, and one or other needs to be high quality. Other quality levels do not change the initial status.

Some remarks help to explain the table and relate it to other approaches that address external validity.

1. If studies on the target population would on their own establish causality in the target population, this is strong, but not infallible, evidence for causation in the target. If there is a study population for which similarity to the target has been established but causation has been ruled out in the study population, then causation in the target population is downgraded to *provisionally established*. (Note that this situation is not covered by the protocol for evaluating external validity advocated by the International Agency for Research on Cancer (IARC); see Sect. 8.1 for further discussion of this point.)
2. Changing the preliminary status of a causal claim obtained from evidence gathered on the target population is more common when that evidence is of lower quality. For instance, a *provisionally established* status may be changed to *established* only in case of established efficacy in the study and established similarity between study and target. The status *arguable*, however, may be changed in case of established efficacy in the study and provisionally established similarity.
3. The GRADE working group also evaluates whether evidence from a study population can be used to draw inferences about the target population. In particular, the GRADE working group considers the case where no evidence directly obtained on the target population is available:

In general, one should not rate down for population differences unless one has compelling reason to think that the biology in the population of interest is so different from that of the population tested that the magnitude of effect will differ substantially. Most often, this will not be the case. [...] The above discussion refers to different human populations, but sometimes the only evidence will be from animal studies, such as rats or primates. In general, we would rate such evidence down two levels for indirectness (Guyatt et al. 2011, pp. 1304–1305)

Hence, the GRADE working group takes similarity of mechanisms to be *established* by default when study and target populations are both human populations. This is problematic because it sets the standard of evidence required for extrapolation too low. In the case of animal studies, one can interpret the default assumption of the GRADE working group as being that the causal claim is *arguable* solely on the basis of causation in animals having been established. Again this is problematic. In our approach, in the absence of evidence of similarity of mechanisms, efficacy in the study population cannot be extrapolated to the target. Hence, even if many high quality RCTs in animals establish efficacy in animals, in the absence of evidence of similarity, nothing can be concluded about efficacy in humans. There is thus a sense in which the approach presented here is more cautious than the GRADE approach to external validity.

4. Causation can be established or ruled out even where no clinical studies on the target are available. This is the case when causation has been established in a study population for which it has been established that it is mechanistically similar to the target population. (This case is captured by the fourth row of Table 7.1, where causation in the target is *speculative*.)
5. Note that, by similarity of mechanisms we mean that any mechanisms in the target population which counteract this mechanism do not mask the effect of the mechanism of action to such an extent that a net correlation in the target

population could not be explained mechanistically (see Sect. 2.3). Consequently, with a mechanism established and some counteracting mechanisms established in the study, a small correlation may be good evidence for causation in the target even if it is not the case that the whole mechanistic structure is similar. After all, this counteracting mechanism would only make the existent correlation smaller in the study than in the target.

7.3 Presenting the Status of a Causal Claim

In presenting the status of a causal claim the following questions need to be addressed, and the status of the causal claim presented after the evaluation of evidence.

Presenting the status of the efficacy claim. The following questions should be addressed:

1. What is the population to which the status applies?
2. What is the intervention or exposure level?
3. What is the outcome and how is it measured?
4. What is the status of the correlation claim? How is this status obtained?
5. What is the status of the general mechanistic claim? How is this status obtained? (See Chap. 6.)
6. What is the status of the efficacy claim?

The following box considers the case where efficacy is extrapolated from one to another population

Presenting the status of the effectiveness claim. The following questions should be addressed:

1. What is the target population to which the status applies?
2. What is the intervention or exposure level in the target?
3. What is the outcome and how is it measured in the target?
4. What is the study population?
5. What is the intervention or exposure level in the study?
6. What is the outcome and how is it measured in the study?
7. What is the status in the study? How is this status obtained?
8. What is the status in the target obtained by evidence directly of the target? How is this status obtained?
9. What is the status of the general mechanistic claim, i.e. that target and study are similar? (See Chap. 6.)
10. What is the overall status of the effectiveness claim?

Standard evidence appraisal systems can be extended to take these considerations into account. For an example of how to incorporate certain aspects of this procedure into a GRADE-style evidence profile, see Sect. 4.6.

References

- Clarke, B., & Russo, F. (2016). Causation in medicine. In J. Marcum (Ed.), *Companion to contemporary philosophy of medicine* (pp. 297–322). Bloomsbury.
- Clarke, B., & Russo, F. (2017). Mechanisms and biomedicine. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy, Chap. 24* (pp. 319–331). London: Routledge.
- 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.
- Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., et al. (2011). Grade guidelines: 1. Introduction grade evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology, 64*(4), 383–394.
- Illari, P. M. (2011). Mechanistic evidence: Disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science, 25*(2), 139–157.
- Joffe, M. (2011). The gap between evidence discovery and actual causal relationships. *Preventive Medicine, 53*(4–5), 246–249.
- Russo, F. (2011). Causal webs in epidemiology. *Paradigmi, XXXIX*(1), 67–98.
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science, 21*(2), 157–170.
- Russo, F., & Williamson, J. (2011). Epistemic causality and evidence-based medicine. *History and Philosophy of the Life Sciences, 33*(4), 563–582.
- Russo, F., & Williamson, J. (2012). EnviroGenomarkers: The interplay between mechanisms and difference making in establishing causal claims. *Medicine Studies: International Journal for the History, Philosophy and Ethics of Medicine & Allied Sciences, 3*, 249–262.
- Williamson, J. (2018). Establishing causal claims in medicine. *International Studies in the Philosophy of Science*, in press. <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.

