

★

## Establishing the teratogenicity of Zika and evaluating causal criteria

★

Jon Williamson

Draft of June 27, 2018

### Abstract

The teratogenicity of the Zika virus was considered established in 2016, and is an interesting case because three different sets of causal criteria were used to assess teratogenicity. This paper appeals to the thesis of Russo and Williamson (2007) to devise an epistemological framework that can be used to compare and evaluate sets of causal criteria. The framework can also be used to decide when enough criteria are satisfied to establish causality. Arguably, the three sets of causal criteria considered here offer only a rudimentary assessment of mechanistic studies, and some suggestions are made as to alternative ways to establish causality.

*Keywords.* Hill Criteria; Shepard Criteria; Russo-Williamson Thesis; Causal criteria; Causal indicators; Zika virus; Teratogenicity

### §1

#### Introduction

In his hugely influential 1965 paper Austin Bradford Hill put forward a set of evidential criteria for establishing a causal claim (Hill, 1965).<sup>1</sup> Hill's criteria are now widely used, particularly when assessing a causal claim in the absence of conclusive evidence provided by randomised controlled trials (RCTs). Note that there are many situations in which RCTs on humans are rarely, if ever, available: for instance, when assessing harms caused by environmental exposure to chemicals—e.g., assessing the carcinogenicity of quinolene—or assessing harms caused by infectious disease—e.g., assessing the teratogenicity of Zika.

While Hill's criteria provide an idea of the kinds of considerations to take into account when assessing a causal claim, such as the dose-response gradient, Hill offers no clear advice as to how the criteria combine to provide an overall evaluation of the causal claim. Moreover, there are sets of causal criteria other than Hill's and it is unclear as to how to compare the relative merits of different sets of criteria.

This paper puts forward a new approach to these problems. It develops an evidential framework that can be used to understand the role of causal criteria in

---

<sup>1</sup>These are indicators of causality, rather than necessary and sufficient conditions for causality (see, e.g., Phillips and Goodman, 2004). However, these indicators are often called 'causal criteria', and we will adopt this entrenched terminology here.

establishing a causal claim and to help decide when causality is established. Such an overarching framework is an important step towards clearer and better methods for assessing causal claims.

The paper is structured as follows. In §2 we see that at least three different sets of causal criteria were used to show that Zika virus causes birth defects: Hill's criteria, Shepard's criteria and a third, new set of criteria. This leads to two questions: How should sets of criteria be evaluated and compared? How should criteria within a single set of criteria weigh against one another to establish causality? In order to address these questions, §3 develops a new evidential framework. In §4, this framework is used to shed some light on the two questions. §5 shows that the framework is amenable to a quantitative Bayesian analysis, if required. Conclusions are drawn in §6.

## §2

### How was teratogenicity established?

Zika virus is a flavivirus, first identified in monkeys in the Zika forest in Uganda in 1947 and in humans in Uganda and Tanzania in 1952. It is spread by mosquitoes of the *Aedes* genus, which also transmit dengue, chikungunya and yellow fever. The first large outbreak was reported on the island of Yap, Micronesia, in 2007, and the virus is now present across tropical regions of Africa, the Americas, Asia and the Pacific. It achieved notoriety in 2015 when an outbreak in Brazil was linked to birth defects.

It is now widely agreed that Zika virus infection during pregnancy has been established to be a cause of abnormal development of the embryo, leading to stillbirths, microcephaly, damage to eyesight and hearing, and Guillain-Barré syndrome. (Other congenital abnormalities are speculated to be effects of Zika, but these links have arguably not been established.) A cause of abnormal physiological development is called 'teratogenic'. The question arises as to how the teratogenicity of Zika virus was established.<sup>2</sup> It turns out that this question poses an interesting problem for medical methodology, because epidemiological studies on their own were insufficient to establish causality—other indicators of causality were required—, and it is not clear how the lists of criteria compare, nor how criteria within a list combine to establish causality.

Three studies are particularly relevant to the question of how the teratogenicity of Zika became established.

Frank et al. (2016) applied Austin Bradford Hill's criteria (Table 1) to the question of whether Zika virus is a cause of microcephaly, and concluded that causality was not established. This study was published in March 2016. The authors argued as follows: there was insufficient evidence of a strong association; evidence of the association was inconsistent but limited; there was no specificity, because other causes of microcephaly were known;<sup>3</sup> temporality was satisfied in individual

---

<sup>2</sup>A causal claim is *established* when the epistemological standard is reached for treating the claim as evidence for other claims. Epistemologists disagree as to what constitutes evidence and when one can treat a proposition as evidence—*established* is used here as a neutral term, not allied to any particular epistemological theory of evidence.

<sup>3</sup>Hill (1965) introduces specificity in the context of a case where the cause is not associated with other diseases. Perhaps he would also view as specific the situation in which the disease is not associated with other causes, which is the situation considered by Frank et al. (2016). Either way, the specificity desideratum is not satisfied in the case of Zika.

Table 1: The criteria of Hill (1965).

Strength	Strength of the observed association
Consistency	Consistency of the observed association
Specificity	A narrowly defined cause and effect (disease), and the cause is not associated with other diseases
Temporality	The putative cause occurs before early stages of the disease
Biological gradient	A dose-response curve
Plausibility	Plausible given the biological knowledge of the day
Coherence	No conflict with the known history and biology of the disease
Experiment	Confirming experimental evidence
Analogy	Similar effects of similar causes

Table 2: The criteria of Shepard (1994).

1	Proven exposure to agent at critical time(s) in prenatal development
2	Consistent findings by two or more epidemiologic studies of high quality
3	Careful delineation of the clinical cases
4	Rare environmental exposure associated with rare defect
5	Teratogenicity in experimental animals
6	The association should make biologic sense
7	Proof in an experimental system that the agent acts in an unaltered state

cases; evidence for biological gradient was apparently not yet available; there was biological plausibility; coherence was satisfied; experimental evidence was limited; there was significant analogical evidence (two other flaviviruses cause congenital brain malformations and other viruses cause microcephaly). Frank et al. (2016) champion the Hill criteria as a means of establishing causality in this case, but at a rather intuitive level: they do not specify which combinations of criteria would be sufficient to establish causality. However, they clearly suggest that, in combination, the evidence was not enough to establish causality in the Zika case. Doshi (2016), in a *BMJ* editorial published in April 2016, also suggests that causality was not established at that time. Broutet et al. (2016), also published in April, concurs.

Rasmussen et al. (2016), on the other hand, applied Shepard’s criteria for establishing teratogenicity (Table 2), and concluded that causality *was* established. Their study was published in May 2016. They argued that criteria 1,3,4 and 6 were met, while the others were not—in particular, epidemiological evidence was limited. Note that Shepard maintained that, to establish teratogenicity, it is necessary to satisfy criteria 1 and 3, and either 2 or 4 (Shepard, 1994, Table 1). Rasmussen et al. (2016) interpret Shepard as saying that such a combination of conditions is *sufficient* to establish teratogenicity, and that since criteria 1, 3 and 4 were satisfied, teratogenicity was established. They also say that all Hill’s criteria except for *experiment* were met, and concluded that teratogenicity was established also on Hill’s account. They attributed the difference of opinion with Frank et al. (2016) to new evidence: two epidemiological studies, an experimental study and a case report. Rasmussen

Table 3: The criteria of Krauer et al. (2017).

---

Temporality
Biological plausibility
Strength of association
Exclusion of alternative explanations
Cessation
Dose-response relationship
Experiments in animals
Analogy
Specificity of association
Consistency

---

et al. (2016) represented the USA-based Centers of Disease Control and Prevention (CDC), an influential organisation, which subsequently announced, ‘Based on rigorous peer-reviewed evaluation of the scientific evidence, CDC and international partners have concluded that Zika virus infection during pregnancy is a cause of microcephaly and other severe brain defects (CDC, 2016, p. 33).’

The third relevant study is that of Krauer et al. (2017), representing the position of the World health Organisation (WHO). This study was accepted for publication in November 2016 and was published in January 2017. The study took the form of a systematic review of the literature and appealed to the list of criteria of Table 3. Krauer et al. (2017) did not define these terms, but several of them are intended to be versions of corresponding criteria of Hill (Broutet et al., 2016). There are three criteria not present in Hill’s list. *Experiments in animals* is an analogue of Shepard’s criterion (5). When assessing whether Zika virus causes congenital brain abnormalities, *exclusion of alternative explanations* considered alternative infections, maternal exposure to alcohol or medication, genetic causes and environmental toxins and heavy metals. *Cessation* considered the reduction in congenital abnormalities following seasonal decline in the vector mosquitoes and following increase in population immunity. The overall conclusion was that Zika virus infection being a cause of congenital brain abnormalities and Guillain-Barré syndrome was the most likely explanation of available evidence. The authors do not explicitly infer this best explanation, nor do they explicitly claim to have established causality. However, they do say that ‘We reached the same conclusion as Rasmussen et al. (2016), but the larger number of studies allowed a more comprehensive and balanced summary of evidence and of evidence gaps’ (Krauer et al., 2017, p. 17), which suggests that they take causality to be established. In support of this suggestion, on 6 September 2016, shortly after this systematic review was submitted for publication (which was on 25 August 2016), the WHO stated that ‘There is scientific consensus that Zika virus is a cause of microcephaly and Guillain-Barré syndrome’ (WHO, 2016).

It is fair to say, then, that the teratogenicity of Zika virus was considered established by the main players in the community mid-to-late 2016. However, three different sets of criteria were applied to assess causality.<sup>4</sup> Given that multiple sets of criteria were invoked and different conclusions were reached, it is not clear whether

---

<sup>4</sup>In addition, in February 2016 a WHO committee also suggested the use of Koch’s postulates (Heymann et al., 2016, p. 720), a view endorsed by Doshi (2016).

the difference in conclusions is attributable to a difference in evidence, the difference in the chosen set of criteria, or the way in which the criteria were applied and weighed against one another. From a methodological point of view, this raises two questions. First, how should sets of criteria be evaluated and compared? Second, how should different criteria within a single set of criteria weigh against one another to establish causality?

### §3

#### An evidential framework

These two questions will be addressed by building an epistemological framework that can be used to integrate and evaluate causal criteria. This framework will appeal to two key observations.

The first is the epistemological thesis of [Russo and Williamson \(2007, §§1–4\)](#). This says that in order to establish a causal claim in medicine, one normally needs to establish two things: that the putative cause and effect are appropriately correlated and that there is some underlying mechanism which can explain instances of the putative effect by appeal to the putative cause and can account for magnitude of the observed correlation. It is not enough to establish a correlation on its own, because, as is well known, a correlation may have one of a number of explanations, only one of which is causation—others include various biases and confounding, for instance. The existence of a mechanism of action distinguishes those correlations that are causal from those that are not. On the other hand, it is not enough to establish the existence of a mechanism on its own, because mechanisms are complex, involving multiple entities, activities and organisational features, and the existence of a mechanism connecting the putative cause to the putative effect does not guarantee a net effect. Moreover, mechanisms can counteract one another, again leading to an absence of a net effect. The existence of an overall correlation distinguishes those mechanisms that underpin genuine causal relationships from those that do not. Thus, one needs to establish the existence of both a correlation and a mechanism.

[Williamson \(2018\)](#) provides a recent detailed statement and a defence of this epistemological thesis. There has been some controversy around a related suggestion, namely that one needs to identify the details of a mechanism in order to establish a causal claim (see, e.g., [Broadbent, 2011](#); [Howick, 2011](#)). However, the Russo-Williamson Thesis only requires establishing the *existence* of a mechanism and the *existence* of a correlation—not the details of the mechanism nor the extent of the correlation ([Illari, 2011](#); [Williamson, 2018, §2.1](#)). The thesis has been supported by analysing medical practice in a number of historical case studies. For example: establishing that the Epstein-Barr virus is a cause of Burkitt’s Lymphoma, and establishing that the human papillomavirus is a cause of cervical cancer ([Clarke, 2011](#)); establishing that smoking causes lung cancer, failing to establish that heavy drinking causes lung cancer, and establishing smoking as a cause of heart disease ([Gillies, 2011](#)). Surveys of present-day research papers also provide some support: for example, [Russo and Williamson \(2011\)](#) argue for the thesis in the practice of autopsy and [Darby and Williamson \(2011\)](#) cite case studies in biomedical imaging.

In sum, the first key component of the epistemological framework is the Russo-Williamson Thesis. While this first observation concerns what needs to be established, the second observation concerns the studies that do the establishing. The observation is that these studies can broadly be divided into two kinds.

On the one hand there are studies that repeatedly measure the putative cause  $A$  and effect  $B$  together, usually in conjunction with other variables that are potential confounders. These studies are often called *clinical studies* or *statistical studies* for assessing whether  $A$  is a cause of  $B$ . There are various subclasses of such studies. In an *experimental study* such as a randomised controlled trial (RCT), the measurements are made after an experimental intervention. If no intervention is performed, the study is an *observational* or *epidemiological study*: a *cohort study* follows a group of people over time; a *case control study* divides the study population into those who have a disease and those who do not and surveys each cohort; a *case series* is a study that tracks patients who received a similar treatment or exposure. An *n-of-1* study consists of repeated measurements of a single individual; other studies measure several individuals. For ease of reference we will use ‘clinical studies’ to refer to all these kinds of study, despite the fact that such studies need not be conducted in the clinical setting.

The second kind of study is a *mechanistic study*. This kind of study investigates features of the mechanism by which  $A$  is hypothesised to cause  $B$ . For example, it might determine whether some further variable  $C$  is an intermediary between  $A$  and  $B$ , or it might investigate the entities, activities or organisational features of the mechanism of action. Mechanistic studies can involve direct manipulation (e.g., *in vitro* experiments), direct observation (e.g., biomedical imaging, autopsy, a case report), confirmed theory (e.g., biochemistry), analogy (e.g., animal experiments) or simulation (e.g., agent-based models). In addition, a clinical study for the claim that  $A$  is a cause of  $C$ , where  $C$  is an intermediate variable on the mechanism from  $A$  to  $B$ , is also a mechanistic study for the claim that  $A$  is a cause of  $B$  because it provides evidence of the details of the mechanism from  $A$  to  $B$ . However, a clinical study for the claim that  $A$  is a cause of  $B$  is not normally a mechanistic study for the claim that  $A$  is a cause of  $B$  because, although it can provide indirect evidence that there exists some mechanism linking  $A$  and  $B$ , it does not normally provide evidence of the structure or features of that mechanism. Similarly, a mechanistic study for the claim that  $A$  is a cause of  $B$  is not normally a clinical study for the claim that  $A$  is a cause of  $B$ , because it does not measure values of  $A$  and  $B$  together. A study will be called a *mixed study* if it is both a clinical study and a mechanistic study—i.e., if it both measures values of  $A$  and  $B$  together and provides evidence of features of the mechanism linking  $A$  and  $B$ . For clarity of exposition and since mixed studies are rare, mixed studies will not feature in the framework developed below—the framework will consider only clinical and mechanistic studies that are not mixed studies. Having grasped the basic framework, it should be clear how to integrate mixed studies into the framework.

We now have the ingredients for the epistemological framework that will be used to integrate and evaluate causal criteria. This framework is based on the evidential relationships portrayed in Fig. 1. The connections between the top three nodes depict the observation that establishing that  $A$  is a cause of  $B$  requires establishing that  $A$  and  $B$  are appropriately correlated and that there exists an appropriate mechanism linking  $A$  and  $B$ . There are two ways of confirming this latter general mechanistic claim. Clinical studies (e.g., high-quality, large RCTs) can confirm the general mechanistic claim (channel  $C_2$  in Fig. 1) if they find a correlation that can best be explained by the general mechanistic claim being true, rather than by bias or confounding, say. Note that that such studies confirm the claim that some mechanism of action *exists* without shedding light of the details of this mechanism. Hence the route of confirmation, labelled  $C_2$ , proceeds directly

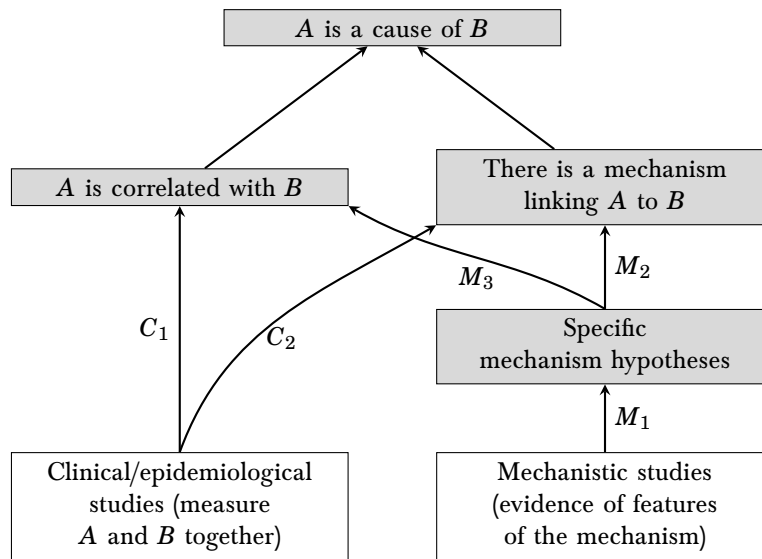


Figure 1: Evidential relationships for establishing a causal claim.

to the general mechanistic claim rather than via a specific mechanism hypothesis. A *specific mechanism hypothesis* is a hypothesis of the form: a particular mechanism with certain features  $F$  can account for the extent of the observed correlation. A specific mechanism hypothesis need not identify all the features of a mechanism of action.

The second way of confirming the general mechanistic claim is by identifying the actual mechanism of action (channel  $M_2$ ), whose features are confirmed by mechanistic studies ( $M_1$ ). In addition, clinical studies provide good evidence of correlation ( $C_1$ ), and, in certain circumstances, an established mechanism of action can also provide good evidence of correlation (channel  $M_3$ ).<sup>5</sup>

#### §4

#### Evaluating causal criteria

We are now in a position to explore the three sets of causal criteria that were used to assess the teratogenicity of Zika virus.

First let us consider Hill's criteria (Table 1). *Strength of association* and *consistency of association* primarily assess clinical studies and they play two roles. First, they are used to infer a correlation (channel  $C_1$ ). Second, a strong, consistent association is also less likely to be spurious, so when these criteria are present, confirmation also flows along channel  $C_2$ . *Specificity*, *temporality*, and *biological gradient* also assess features of clinical studies and support the existence of a mechanism without elucidating the details of the mechanism of action (channel  $C_2$ ). Similarly, it is clear from Hill's discussion of his criteria that the *experiment* criterion is satisfied when the clinical study is an experimental study, rather than by experiments that

<sup>5</sup>See Williamson (2018, §2.2.3) for one kind of example of  $M_3$ -channel confirmation and Pearl and Bareinboim (2014) for another.

Table 4: A classification of the three sets of criteria according to the main evidential channels along which each criterion operates.

	Hill	Shepard	Krauer et al.
<i>C</i> -channel	Strength	1	Strength of association
	Consistency	2	Exclusion of alternatives
	Specificity	3	Specificity
	Temporality	4	Temporality
	Biological gradient		Dose-response
	Experiment		Cessation
	Analogy	5	Animal experiments
<i>M</i> -channel	Analogy	7	Analogy
	Plausibility	6	Biological plausibility
	Coherence		

shed light on the mechanism of action (Hill, 1965, p. 298). Thus this criterion also operates along channel  $C_2$ . *Plausibility* and *coherence*, on the other hand, require that the existence of a mechanism of action should fit with evidence of the relevant mechanisms—i.e., these assess channel  $M_2$ . For Hill, *Analogy* operates by increasing the evidence base by considering results from other situations that are known or suspected to be mechanistically similar to the case in hand: ‘With the effects of thalidomide and rubella before us, we would surely be ready to accept slighter but similar evidence with another drug or another viral disease in pregnancy’ (Hill, 1965, p. 11). Thus *Analogy* primarily operates by lowering the burden of proof in the  $C_1$  and  $C_2$  channels; it does not provide details of the mechanism of action.

Let us turn to Shepard’s criteria (Table 2). Criteria 1-4 pertain to clinical studies. (1) *proven exposure* and (3) *careful delineation* primarily help to establish a correlation (channel  $C_1$ ); (2) *consistent findings* and (4) *rare-exposure-rare-defect* operate via both  $C_1$  and  $C_2$ . (5) *teratogenicity in experimental animals* and (7) *proof in an experimental system* provide evidence that there is a robust correlation and a mechanism of action without shedding light on any specific mechanism hypothesis (i.e.,  $C_1$  and  $C_2$ ). (6) *biologic sense* concerns the fit between the existence of a mechanism and confirmed mechanism hypotheses and so evaluates  $M_2$ . Shepard’s necessary condition can now be seen to imply that in order to establish causality it is essential that there is support along both the  $C_1$  and  $C_2$  channels.

Finally, consider the criteria of Krauer et al. (2017), listed in Table 3. As in the case of Hill’s criteria; *temporality*, *specificity* and *dose-response* relationship concern  $C_2$ ; *biological plausibility*,  $M_2$ ; *strength of association*,  $C_1$  and  $C_2$ . *Animal experiments* and *analogy* concern  $C_1$  and  $C_2$ . *Cessation* confirms the existence of a correlation but also confirms the existence of a mechanism without shedding light on features of the mechanism, and so concerns  $C_1$  and  $C_2$ . Finally, *exclusion of alternative explanations* boosts confirmation along the  $C_2$  channel, because, if alternative explanations really are excluded, the correlation can only be attributable to an underlying mechanism of action.

Having seen roughly how the different causal criteria fit into our evidential scheme, we can begin to evaluate the criteria. The first thing to note is that most of the criteria focus on clinical studies and the  $C_1$  and  $C_2$  channels (see Table 4). This



focus can be explained as follows. Only very rarely can one establish the existence of a correlation solely by way of inferring the correlation from the mechanism of action (channel  $M_3$ ). The only other way is an inference from clinical studies ( $C_1$ ). Therefore, clinical studies are almost always required to establish causation. Furthermore, in certain cases clinical studies suffice to establish causation. If there are sufficiently many independently conducted RCTs of sufficient quality that observe a sufficiently large correlation, and certain other explanations of the correlation are ruled out, then one can infer the existence of a mechanism along the  $C_2$  channel (Williamson, 2018, §2.2.1). Therefore—if one is lucky—it can suffice to consider clinical studies and the  $C_1$  and  $C_2$  channels. This explains the focus on the  $C$  channels.

Normally, however, one is not so lucky. The case of establishing the teratogenicity of Zika exemplifies the typical situation: epidemiological studies on humans did not suffice to establish causality and other evidence had to be considered. However, none of the three sets of criteria offer more than a very rudimentary treatment of the  $M$  channels. There is a reason for this. Mechanistic studies are very diverse, as are the specific mechanism hypotheses that they inform, and it is hard to say anything that is general enough to take the form of a domain-independent criterion that might feature in one of the above lists of criteria, yet specific enough not to be hopelessly vague. We are left with vague criteria such as ‘plausibility’, ‘coherence’ and ‘biologic sense’.

There are various ways in which one might respond to this problem of the inadequate treatment of the  $M$  channels. One response is to move from the general to the particular: in particular domains, one can be more informative about the  $M$  channels. For example, the International Agency for Research on Cancer (IARC) has put forward ten key characteristics of carcinogenicity that are used to suggest and assess specific mechanism hypotheses (Smith et al., 2016).

A second response is to apply the quantitative apparatus of Bayesianism to determine the probability of the causal claim conditional on the available clinical and mechanistic studies. Landes et al. (2018) provide one such Bayesian analysis, though determining the parameters of their Bayesian model may be challenging in practice. A somewhat simpler Bayesian analysis is sketched in the next section.

A third response to the problem of the  $M$  channels is to apply a set of qualitative heuristics that fit the evidential relationships set out in Fig. 1. This is the approach of the *EBM+* methodology of Parkkinen et al. (2018). According to this methodology, clinical studies are assessed using standard methods such as the GRADE system (Guyatt et al., 2011), and there is also a parallel stream of assessment corresponding to the  $M$  channels, which proceeds as follows. First, specific mechanism hypotheses are formulated, and next these are assessed by systematically searching for relevant mechanistic studies and then evaluating the evidence for key features of each hypothesised mechanism of action (the  $M_1$  channel). Parkkinen et al. (2018) then provide heuristics for combining this evaluation with the assessment of clinical studies, in order to assess the general mechanistic hypothesis that there is some mechanism of action (i.e., to assess the  $C_2$  and  $M_2$  channels). Finally, they provide further heuristics for assessing the causal claim itself, given the status of the correlation claim and the status of the claim that there is a mechanism of action. This yields an overall evaluation of the status of the causal claim, on the basis of both clinical and mechanistic studies.

As to which response is appropriate will depend on the particular circumstances. In the Zika case, the *EBM+* approach may well be the most fruitful, since Shepard’s

domain-specific criteria of Table 2 do not shed much light on the  $M$  channels, and a quantitative Bayesian analysis may be hard to carry out.

Alternatively, if an approach based on one or other set of causal criteria is chosen after all, the evidential framework represented in Fig. 1 can help to weigh criteria against one another in order to decide whether causality is established. In order for causality to be established, one needs the following combination of criteria. First,  $C_1$ -channel criteria and any  $M_3$ -criteria need to be satisfied to extent that the  $C_1$  and  $M_3$  channels jointly establish the existence of a correlation. Second,  $C_2$  criteria and  $M_1$  and  $M_2$  criteria need to be satisfied to the extent that the  $C_2$  and  $M_2$  channels jointly establish the existence of a mechanism.

## §5

### A Bayesian analysis

This section sketches a simple Bayesian analysis that conforms to the evidential relationships of Fig. 1. This serves several purposes. First, it shows that the general epistemological framework of §3 is compatible with a formal Bayesian analysis, which lends some support to the framework. Second, it illustrates one way to respond to the problem of the inadequate treatment of the  $M$ -channels by the sets of causal criteria, noted above. Third, this quantitative approach can be used to show which evidential channels in Fig. 1 are particularly important in certain circumstances.

In what follows,  $C$  refers to the clinical study evidence;  $M$  to the mechanistic study evidence;  $c$  to the claim that  $A$  and  $B$  are appropriately correlated;  $m$  to the claim that  $A$  and  $B$  are appropriately mechanistically connected;  $h_1, \dots, h_k$  to the specific mechanism hypotheses that have been proposed as the mechanism of action (which we shall take to be mutually exclusive); and  $h_0$  to the catch-all hypothesis, i.e., the claim that none of  $h_1, \dots, h_k$  can be responsible for the observed correlation. Then we have:

$$P(A \text{ causes } B | CM) = P(cm | CM) \quad (1)$$

$$= \sum_{i=0}^k P(c|h_i C) P(m|h_i C) P(h_i | M) \quad (2)$$

Equation 1 captures the thesis of Russo and Williamson (2007). Equation 2 follows if we assume that Fig. 2 represents the conditional independence structure of the evidential relationships (motivated by Fig. 1). An undirected graph such as Fig. 2 represents conditional independence relationships by the following rule: if set  $Z$  of variables separates set  $X$  from set  $Y$ , then  $X$  is probabilistically independent of  $Y$  conditional on  $Z$ , which is often written as  $X \perp\!\!\!\perp Y | Z$ . For example,  $C$  and  $h$  separate  $M$  and  $m$  from  $c$ , so  $c \perp\!\!\!\perp m, M | C, h$ . Here  $h$  is a variable that takes the hypotheses  $h_0, \dots, h_k$  as values.  $c$  and  $m$  can be construed as binary variables.<sup>6</sup>

In a Bayesian analysis, it can be difficult to determine probabilities that involve a catch-all hypothesis such as  $h_0$ . Here, however, there are circumstances in which this problem can be mitigated. When the clinical studies are inconclusive on their own (e.g., when they are observational studies), and when enough is known about

<sup>6</sup>That Equation 2 holds can be seen by viewing it as the factorisation corresponding to a Bayesian network representation of the the same conditional independence structure.

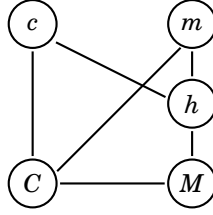


Figure 2: The conditional independence structure of evidential relationships.

the domain and sufficiently many specific mechanism hypotheses have been put forward,  $P(m | h_0 C)$  will be small: should each mechanism hypothesis  $h_1, \dots, h_k$  be ruled out, it will be unlikely that there is in fact any mechanism of action. Moreover, in such a situation,  $h_0$  will also disconfirm  $C$ , and  $P(c | h_0 C)$  will not be large. In which case, the contribution to the sum in Equation 2 made by the catch-all  $h_0$  may well be negligible. In addition,  $P(m | h_i C) = 1$  for  $i = 1, \dots, k$  (if one of the specific mechanism hypotheses is true then trivially there is some appropriate mechanism). Then,

$$P(A \text{ causes } B | CM) \approx \sum_{i=1}^k P(c | h_i C) P(h_i | M).$$

The first term in the product on the right-hand side evaluates the contribution of the  $C_1$  and  $M_3$  channels. The second term quantifies the  $M_1$  channel. The first term can often be further simplified: when there is a reasonable amount of clinical study evidence available, knowing the specific mechanism of action will not tell us much more about the existence of a correlation, so  $P(c | h_i C) \approx P(c | C)$ . In that case,

$$P(A \text{ causes } B | CM) \approx \sum_{i=1}^k P(c | C) P(h_i | M).$$

Under these conditions, then, the  $C_1$  and  $M_1$  channels are paramount. Crucially, the  $M_1$  channel is as important as the  $C_1$  channel. This fact points to limitations of current causal criteria, and indeed much of current practice in establishing causality in the health sciences, which tends to underplay the role of mechanistic studies.

## §6

### Conclusion

A problem is posed by the use of multiple sets of criteria to establish a causal claim such as the teratogenicity of Zika. It is not clear how to evaluate and compare sets of criteria, nor how to decide whether, for a given set of criteria, the evidence suffices to establish causality. The epistemological framework developed in this paper is intended to address this problem. By structuring the criteria according to the evidential channels of Fig. 1 that they assess, one can evaluate how well a set of criteria assesses the channels and evaluate whether, when a subset of criteria are satisfied, those criteria could be enough to establish causality.

Unfortunately, none of the three sets of criteria considered here can be said to assess mechanistic studies in a comprehensive way. To do that, we need to move away from very general lists towards more domain-specific, structured criteria, as

IARC have done, or to perform a quantitative Bayesian analysis such as is provided in §5, or to apply heuristics (e.g., the EBM+ methods) that better fit the evidential relationships of Fig. 1.

Note that the approach developed in this paper differs markedly from the analysis of Bird (2011), who also puts forward a framework for understanding and systematising the Hill criteria. Bird treats causal inference in the absence of experimental studies as a process of elimination: if one can eliminate the possibility that  $B$  causes  $A$ , the possibility that  $A$  and  $B$  have a common cause, and the possibility that there is no causal relation between  $A$  and  $B$ , then one can infer that  $A$  is a cause of  $B$ . One cannot argue with this inference from a logical point of view. However, Bird’s scheme does not consider what are perhaps the primary grounds for eliminating these alternative possibilities in the absence of experimental clinical studies, namely mechanistic studies. The epistemological framework developed here puts clinical and mechanistic studies centre stage and attempts to analyse their mutual interactions.

Howick et al. (2009) developed another framework for understanding Hill’s criteria, and mechanistic studies feature in their account. They divide criteria into ‘direct evidence’ (which acts via the  $C$  channels of the framework developed here), ‘mechanistic evidence’ (the  $M_1$  channel) and ‘parallel evidence’ (which deals with coherence amongst studies). The account presented here can be viewed as providing an explanation of why and how direct evidence and mechanistic evidence complement one another.

In sum, the epistemological framework of this paper can be viewed as compatible with the proposals of Bird (2011) and Howick et al. (2009). But it arguably goes further than their accounts, in that it can explain the relative importance of causal criteria and it can be used to compare sets of criteria.

### **Acknowledgements**

This work was undertaken as a part of the research projects *Evaluating evidence in medicine*, funded by the UK Arts and Humanities Research Council, and *Grading evidence of mechanisms in physics and biology*, funded by the Leverhulme Trust. I am very grateful to Juergen Landes, Michael Wilde, the editors and two anonymous referees for helpful comments.

---

## Bibliography

---

- Bird, A. (2011). The epistemological function of Hill's criteria. *Preventive Medicine*, 53(4):242–245. Special Section: Epidemiology, Risk, and Causation.
- Broadbent, A. (2011). Inferring causation in epidemiology: mechanisms, black boxes, and contrasts. In Illari, P. M., Russo, F., and Williamson, J., editors, *Causality in the sciences*, pages 45–69. Oxford University Press, Oxford.
- Broutet, N., Krauer, F., Riesen, M., Khalakdina, A., Almiron, M., Aldighieri, S., Espinal, M., Low, N., and Dye, C. (2016). Zika virus as a cause of neurologic disorders. *New England Journal of Medicine*, 374(16):1506–1509. PMID: 26959308.
- CDC (2016). Zika virus: Information for clinicians. Published online 23 June 2016. Centers for Disease Control and Prevention, <http://www.cdc.gov/zika/comm-resources/clinicianppt.pptx>.
- Clarke, B. (2011). *Causality in medicine with particular reference to the viral causation of cancers*. PhD thesis, Department of Science and Technology Studies, University College London, London.
- Darby, G. and Williamson, J. (2011). Imaging technology and the philosophy of causality. *Philosophy & Technology*, 24(2):115–136.
- Doshi, P. (2016). Convicting Zika. *BMJ*, 353.
- Frank, C., Faber, M., and Stark, K. (2016). Causal or not: applying the Bradford Hill aspects of evidence to the association between Zika virus and microcephaly. *EMBO Molecular Medicine*, 8(4):305–307.
- Gillies, D. A. (2011). The Russo-Williamson thesis and the question of whether smoking causes heart disease. In Illari, P. M., Russo, F., and Williamson, J., editors, *Causality in the Sciences*, pages 110–125. Oxford University Press, Oxford.
- Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., deBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., and Schünemann, H. J. (2011). GRADE guidelines: 1. introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64(4):383–394.
- Heymann, D. L., Hodgson, A., Sall, A. A., Freedman, D. O., Staples, J. E., Althabe, F., Baruah, K., Mahmud, G., Kandun, N., Vasconcelos, P. F. C., Bino, S., and Menon, K. U. (2016). Zika virus and microcephaly: why is this situation a PHEIC? *The Lancet*, 387(10020):719–721.
- Hill, A. B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, 58:295–300.
- Howick, J. (2011). Exposing the vanities—and a qualified defence—of mechanistic evidence in clinical decision-making. *Philosophy of Science*, 78(5):926–940. Proceedings of the Biennial PSA 2010.
- Howick, J., Glasziou, P., and Aronson, J. K. (2009). The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute? *Journal of the Royal Society of Medicine*, 102:186–194.
- Illari, P. M. (2011). Disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science*, 25(2):139–157.
- Krauer, F., Riesen, M., Reveiz, L., Oladapo, O. T., Martínez-Vega, R., Porgo, T. V., Haefliger, A., Broutet, N. J., Low, N., and Group, W. Z. C. W. (2017). Zika virus infection as a cause of congenital brain abnormalities and Guillain-Barré

- syndrome: Systematic review. *PLOS Medicine*, 14(1):1–27.
- Landes, J., Osimani, B., and Poellinger, R. (2018). Epistemology of causal inference in pharmacology: Towards a framework for the assessment of harms. *European Journal for Philosophy of Science*, 8(1):3–49.
- Parkkinen, V.-P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M. P., Norrell, C., Russo, F., Shaw, B., and Williamson, J. (2018). *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*. Springer Briefs. Springer.
- Pearl, J. and Bareinboim, E. (2014). External validity: from do-calculus to transportability across populations. *Statistical Science*, 29(4):579–595.
- Phillips, C. V. and Goodman, K. J. (2004). The missed lessons of sir austin bradford hill. *Epidemiologic Perspectives & Innovations*, 1(1):3.
- Rasmussen, S. A., Jamieson, D. J., Honein, M. A., and Petersen, L. R. (2016). Zika virus and birth defects—reviewing the evidence for causality. *New England Journal of Medicine*, 374(20):1981–1987.
- Russo, F. and Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21(2):157–170.
- Russo, F. and Williamson, J. (2011). Generic versus single-case causality: the case of autopsy. *European Journal for Philosophy of Science*, 1(1):47–69.
- Shepard, T. H. (1994). “Proof” of human teratogenicity. *Teratology*, 50(2):97–98.
- Smith, M. T., Guyton, K. Z., Gibbons, C. F., Fritz, J. M., Portier, C. J., Rusyn, I., DeMarini, D. M., Caldwell, J. C., Kavlock, R. J., Lambert, P. F., Hecht, S. S., Bucher, J. R., Stewart, B. W., Baan, R. A., Coglianò, V. J., and Straif, K. (2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. *Environmental Health Perspectives*, 124:713–21.
- WHO (2016). Zika virus: fact sheet. Published online 6 September 2016. World Health Organisation, <http://www.who.int/mediacentre/factsheets/zika/en/>.
- Williamson, J. (2018). Establishing causal claims in medicine. *International Studies in the Philosophy of Science*, in press.