Bayesian Analysis (2016) **11**, Number 3, pp. 725–752

From Statistical Evidence to Evidence of Causality

A. Philip Dawid*, Monica Musio†, and Stephen E. Fienberg‡

Abstract. While statisticians and quantitative social scientists typically study the "effects of causes" (EoC), Lawyers and the Courts are more concerned with understanding the "causes of effects" (CoE). EoC can be addressed using experimental design and statistical analysis, but it is less clear how to incorporate statistical or epidemiological evidence into CoE reasoning, as might be required for a case at Law. Some form of counterfactual reasoning, such as the "potential outcomes" approach championed by Rubin, appears unavoidable, but this typically yields "answers" that are sensitive to arbitrary and untestable assumptions. We must therefore recognise that a CoE question simply might not have a well-determined answer. It is nevertheless possible to use statistical data to set bounds within which any answer must lie. With less than perfect data these bounds will themselves be uncertain, leading to a compounding of different kinds of uncertainty. Still further care is required in the presence of possible confounding factors. In addition, even identifying the relevant "counterfactual contrast" may be a matter of Policy as much as of Science. Defining the question is as non-trivial a task as finding a route towards an answer.

This paper develops some technical elaborations of these philosophical points from a personalist Bayesian perspective, and illustrates them with a Bayesian analysis of a case study in child protection.

Keywords: benfluorex, causes of effects, counterfactual, child protection, effects of causes, Fréchet bound, potential outcome, probability of causation.

1 Introduction

One function of a Court of Law is to attempt to assign responsibility or blame for some undesirable outcome. In many such cases there will be relevant testimony about statistical or epidemiological evidence arising from studies done on specialized populations, but this evidence addresses the main issue only indirectly, at best. It has until now been unclear how to use such evidence to focus on the issue at hand which involves specific individuals experiencing the undesirable outcome. Although there is a considerable literature on certain aspects of this problem—see, for example, Green et al. (2011), which aims to assist US judges in managing cases involving complex scientific and technical evidence—we consider that there are important logical subtleties that have not as yet been accorded the appreciation they warrant. Here we show that, even in the (very rare) case that we have the best possible and most extensive data on the "effects of

*University of Cambridge, apd@statslab.cam.ac.uk
†University of Cagliari, mmusio@unica.it
‡Carnegie Mellon University, fienberg@stat.cmu.edu

 DOI: 10.1214/15-BA968

causes," and can accept certain very strong but necessary conditions, there will still remain irreducible uncertainty—which we can express as interval bounds—about the relevant "probability of causation." With less than fully perfect data, this interval uncertainty will be further compounded by statistical uncertainty. Such multiple forms of uncertainty raise subtle issues of interpretation and presentation.

The structure of the paper is as follows. In Section 2 we consider a high-profile case where serious side-effects led to the withdrawal of a drug from the market, and, in turn, to litigation against the manufacturer. Since the evidence in this litigation has yet to be presented formally at trial in court, we consider how general evidence of incidence of effects might or might not be relevant to a hypothetical tort action in which an affected patient sues the manufacturer for damages, and we relate this to the distinction we draw in Section 3 between "effects of causes" and "causes of effects." After a brief consideration of inference from statistical data about effects of causes in Section 4, the remainder of the paper focuses on inference about causes of effects, based on a "probability of causation" defined using counterfactual logic. Although this probability is typically impossible to pinpoint on the basis of epidemiological data, however extensive, in Section 5 we give bounds between which it must lie—bounds which, however, will themselves be subject to statistical uncertainty, which we discuss in Section 6. In Section 7 and Section 8 we illustrate our theory with a new analysis of a case study in child protection. Section 9 offers some concluding remarks.

Our approach is Bayesian in several respects. Least importantly, the data-analysis of our case study is handled using Bayesian prior-to-posterior inference, effected by means of a Gibbs sampler implemented in the WinBUGS © software (Lunn et al., 2012). More significantly, our interpretation of probability—for example, the "probability of causation" in an individual case—is fundamentally personalist. Moreover, we consider carefully the relationship between such personalist judgments and the empirical evidence which informs them, making use of exchangeability considerations to relate personal probabilities to frequencies observed in data. We also address the limitations of this approach when aspects of our desired inference cannot, even in principle, be informed by empirical data.

2 Epidemiological Evidence in Litigation

2.1 Epidemiological Background

The drug Mediator, also known as benfluorex, was for many years marketed as an anti-diabetic drug. It was also widely used off-label as an appetite suppressant. In November 2010, however, following the publication of a popular book by Irène Frachon (2010), the French Health Agency CNAM announced its finding that around 500 deaths in France over a 30 year period could be attributed to Mediator—see also Hill (2011). This was based on extrapolation of results in two scientific studies, published at about the same time, focusing on the effects of benfluorex on valvular heart disease. Frachon et al. (2010) showed a significantly higher prevalence of unexplained valvular heart disease in patients taking benfluorex, as compared to controls. Weill et al. (2010) examined

the records of over a million diabetic patients in a cohort study, and reported a higher hospitalisation rate for valvular heart disease in benfluorex takers.

2.2 Litigation

As the news about Mediator reverberated through the media, the French authorities withdrew the drug from sale. At the same time, hundreds of individuals jointly filed a criminal lawsuit against the manufacturer of Mediator, the French pharmaceutical giant Les Laboratoires Servier. The trial has been under way since May 2012, with initial aspects focused on whether the company was guilty of misconduct. At the time of preparation of this article, the issue of whether Mediator was in fact the cause of the heart disease in any of those who brought the lawsuit had yet to be addressed, and no expert scientific testimony had been presented to the court. As of September 2014, however, the company has agreed to compensate over 350 individual plaintiffs.[1]

In the US benfluorex was removed from the marketplace in the 1990s. The banning in 1997 of a related drug, Redux, led to a $12 billion settlement, following a class action by thousands of individuals (Anon, 17 November 2010). Thus considerable attention both in France and elsewhere is focused on the case against Servier.

2.3 Scientific Results

The matched case-control study of Frachon et al. (2010) involved 27 cases of valvular heart disease and 54 controls. Investigators determined whether the patients had or had not used benfluorex.

We display the core data in Table 1. The face-value odds ratio in this table is $(19 \times 51)/(3 \times 8) = 40.1$, but this could be misleading because of confounding factors.[2] A logistic regression analysis reported by Frachon et al. (2010) adjusted for body mass index, diabetes and dexfenfluramine use, and reduced the odds ratio to 17.1 (95% CI 3.5 to 83.0), a value which is still a large and highly significant measure of positive association between benfluorex and valvular heart disease. In the same direction, Weill et al. (2010) computed a risk ratio (though with relatively crude adjustments) of 3.1 (95% CI 2.4 to 4.0).

Benfluorex Use	Cases	Controls	Totals
Yes	19	3	22
No	8	51	59
Totals	27	54	81

Table 1: Raw results from case-control study linking benfluorex and valvular heart disease. Source: Frachon et al. (2010).

[1] http://www.lejdd.fr/Societe/Sante/Mediator-350-victimes-indemnisees-685286#new-reactions

[2] See Holland and Rubin (1988) for when it is necessary and appropriate to adjust for covariate information in such a study.

Robustness of Odds Ratio

While the risk ratio may be a more relevant and incisive measure of the strength of an effect than the odds ratio (and it will feature importantly in our analysis of Section 4 below), it faces a very serious problem: it is simply not possible to compute it from a retrospective study, such as that of Frachon et al. (2010). In contrast, the odds ratio, whether raw or adjusted via a logistic regression, has the important property that it is simultaneously a meaningful measure of association and computable from retrospective data (Altham, 1970; Bishop et al., 1975; Farewell, 1979). Furthermore, under suitable adjustment for covariates it will be a good approximation to the risk ratio when the outcome is rare.

2.4 Toxic Tort—A Hypothetical Case

Now consider a (currently purely hypothetical) case that might be brought on the basis of these scientific reports. A woman with unexplained valvular heart disease sues the manufacturer of benfluorex, claiming that it was this that caused her illness. An epidemiologist, testifying as expert witness for plaintiff, claims that, on the evidence of Dr. Frachon's and Dr. Weill's studies, the medication can cause valvular heart disease. Citing Nicot et al. (2011), who argued that "the probabilistic information, derived from the available epidemiological studies, needs to be considered as part of evidence to establish or refute a causal link between benfluorex and valvular disease for a given patient," this witness goes on to claim that this is evidence for a causal link in the current case. The defendants in turn proffer their expert, who testifies that in the manufacturer's clinical trials there was no evidence of such a side effect. How should the court rule?

The court needs to decide on the cause of this woman's heart disease. But the plaintiff's expert addresses something different, the general scientific question "Can benfluorex be shown to cause heart disease?" For an epidemiologist, the evidence for this would ideally be captured by the risk ratio, though, as we have seen, for the Frachon data we would have to be satisfied with the adjusted odds ratio instead. But even if we had perfect and unassailable statistical evidence in support of this general scientific hypothesis, that would still be only very indirectly relevant to the individual case at issue. We shall see below that the relationship between such a generalisation and the specific issue before the court is extremely subtle. For an extended discussion of the legal nexus of individual causation and the "causes of effects," see Dawid et al. (2014).

3 Causes of Effects and Effects of Causes

One might be tempted to assume that the "effects of causes" (henceforth EoC) and the "causes of effects" (CoE) are related probabilistically via Bayes's theorem. After all, this was how Laplace (1986) introduced the topic: "If an event can be produced by a number n of different causes, the probabilities of these causes given the event are to each other as the probabilities of the event given the causes,..." Later authors recognized the issue to be more complex.

John Stuart Mill distinguished between inferences about effects of causes and about causes of effects, and remarked "... as a general rule, the effects of causes are far more accessible to our study than the causes of effects..." (Mill, 1843, Book 3, Chapter 10, Section 8). Although a similar distinction has sometimes been expressed in statistical contexts (see, e.g. Holland (1986)), Mill's associated warning has largely gone unheeded. We consider that it deserves more careful attention. Though evidently related in some way, problems of CoE are distinct from problems of EoC; indeed, as Mill understood, they are considerably more subtle and difficult to handle.

In this article, which builds on and extends Dawid (2011) and Dawid et al. (2014), we attempt to delineate both the differences and the connexions between these two distinct inferential enterprises. An understanding of these issues will clearly be crucial if generic retrospective observational EoC evidence, such as that of the Frachon et al. (2010) study, is to be brought to bear on an individual CoE case, such as the toxic tort case of Section 2.4. In particular we shall consider the possibilities of using statistical evidence to inform CoE inferences.

Recently Gelman and Imbens (2013) have addressed what they term the distinction between "forward causal inference" and "reverse causal questions." Although this superficially appears similar to the distinction we draw between EoC and CoE, there are some important differences. Our CoE scenario involves an individual case, where the outcome ("effect") is known, and a clearly identified putative cause is being queried. Their "reverse causal question," on the other hand, is "Why?:" Just what, from among many possibilities, might be "the cause" (or the causes) of a general effect? They use a model checking approach to the "effect of causes" framework to explore "Why?," searching for variables that might not yet even be in the model. Developing common ground between their approach and ours could be valuable, but is beyond the scope of the present paper.

3.1 Aspirin Trial

As a simple concrete example, we contrast the following two questions:

Effects of Causes (EoC) Ann has a headache. She is wondering whether to take aspirin. Would that cause her headache to disappear (within, say, 30 minutes)?

Causes of Effects (CoE) Ann had a headache and took aspirin. Her headache went away after 30 minutes. Was that caused by the aspirin?

Note that—in a departure from previous related treatments—in both questions we have separated out the rôles of the subject ("Ann"), on whom we have some information, and the questioner or analyst (henceforth "I"), who wants to interpret that information: these could be the same individual, but need not be. The relevant uncertainty about the answers to the above queries is my own personal uncertainty, informed by relevant data. This can be represented by a subjective probability distribution, P_A, say, over attributes of Ann. Our distinction between the subject and the analyst is somewhat analogous to the situation in court, where we distinguish between a witness, who supplies evidence

(e.g. on epidemiology), and the trier of fact, be it a judge or a jury, who has to assess the uncertainty to associate with the question of ultimate legal interest: the cause of the effect.

What might be relevant data in the present instance, and how might such data be applied? For illustration, we suppose that our data arise from a well-conducted (large, prospective, randomised, double-blind,...) comparative clinical trial in which we have observed the following recovery rates:

$$\Pr(R = 1 \,|\, E = 1) \;=\; 30\% \tag{1}$$
$$\Pr(R = 1 \,|\, E = 0) \;=\; 12\% \tag{2}$$

where $E = 1$ [resp., 0] denotes "exposure to" (= treatment with) aspirin [resp., no aspirin], and $R = 1$ [resp., 0] denotes that the headache does [resp., does not] disappear (within 30 minutes). Here and throughout, we use $\Pr(\cdot)$ to denote probabilities (henceforth termed *chances*) underlying a population data-generating process.

4 Statistical Evidence for EoC

Ann has a headache. She is wondering whether to take aspirin. Would that cause her headache to disappear (within, say, 30 minutes)?

Most of classical statistical experimental design and inference is geared to elucidating the effects of causes, and much careful attention over many years has gone into clarifying and improving methods for doing this, for example, by the use of randomised comparative experiments (Fisher, 1935; Hill, 1951) to control for potential confounding factors. Even when emphasis is specifically targeted on statistical causality (Rubin, 1974; Holland and Rubin, 1988; Pearl, 2009) this still mostly addresses EoC problems, albeit in observational rather than experimental settings.

In order to highlight the major issue, we confine attention here to data from a large study, such as the aspirin trial of Section 3.1, that can be regarded as supporting genuinely causal inferences.[3] Suppose moreover that I regard Ann as being exchangeable with the patients in the trial. Then I can identify my personal probabilities, P_A, for attributes of Ann with the corresponding frequencies, Pr, observed in the data. In particular, if Ann takes aspirin I can expect her headache to disappear within 30 minutes with probability 30%, but with probability only 12% if she does not. If I myself am Ann, then (other things being equal) taking the aspirin is my preferred option.

In this case, the EoC causal inference is based on a simple contrast between the two "prospective" conditional probabilities, $\Pr(R = 1 \,|\, E = 1)$ and $\Pr(R = 1 \,|\, E = 0)$. In particular, the information needed for making EoC causal inferences—and so for guiding future decisions—is subsumed in the conditional probability distribution of the response R given exposure E. In more complex situations we may have to make various modifications, e.g. adjustment for covariates, but the essential point remains that purely

[3] Some considerations relevant to the possibilities for causal inference in various data-collection settings can be found in Dawid (2011).

probabilistic knowledge, properly conditioned on known facts, is sufficient to address EoC-type questions.

5 Statistical Evidence for CoE

Ann had a headache and took aspirin. Her headache went away after 30 minutes. Was that caused by the aspirin?

5.1 How to Understand "Causes of Effects"?

Addressing a CoE-type question is much more problematic—indeed, even to formulate the question clearly is a nontrivial enterprise. We can no longer base our approach purely on the probability distribution of E and R conditioned on known facts, since we know the values of both variables ($E = 1$, $R = 1$), and after conditioning on that knowledge there is no probabilistic uncertainty left to work with.

One possible approach, popular in statistical circles, is based on the concept of the "counterfactual contrast," which in turn rests on the introduction of "potential responses" (Rubin, 1974). We proceed by splitting the response variable R into two variables, R_0 and R_1, where we conceive of R_1 [resp., R_0] as a potential value of R, that will eventuate if in fact $E = 1$ [resp., 0]. Both these potential responses are regarded as existing prior to the determination of E. We thus now need to model the three variables (E, R_0, R_1) together, rather than (as previously) just the two variables (E, R).[4]

We might now cast the CoE question as enquiring about the relationship between R_0 and R_1. Thus "$R_1 = 1, R_0 = 0$" describes the situation where Ann's headache disappears if she takes the aspirin, but does not if she does not—a state of affairs that might reasonably be described as the disappearance of Ann's headache being *caused* by taking the aspirin. In particular, if Ann has taken the aspirin and her headache disappeared (thus $R_1 = 1$), these two events can be regarded as causally connected just in the case that $R_0 = 0$.

5.2 Science and Policy

Although we shall follow through with the above formulation in the remainder of this article, we here turn aside to consider an objection to it: it simply might not be appropriate to regard, as the "counterfactual foil" to the factual response (R_1), what would have happened (R_0) if the exposure had not occurred ($E = 0$) but all other prior circumstances were the same. For example, there has been a series of legal cases in which various administrations have sued tobacco companies on the basis that they had not properly informed the public of the dangers of smoking when they first had that evidence, and should therefore be liable for the increased costs that fell on health services due to that act of omission. But it could be argued that, since smokers tend to die earlier than non-smokers, encouraging (or at least not discouraging) smoking would in

[4]The observed response R is determined by these three variables as $R = R_E$.

fact reduce the total burden on the health services. Such an attempted defence has, however, usually been ruled inadmissible. Instead, as a matter of policy, the relevant counterfactual comparator is taken to be a hypothetical universe in which every one lives just as long as they do in fact, but they are healthier because they smoke less. Here we see Science and Policy as inextricably intertwined in formulating the appropriate CoE question. And the conceptual and implementational difficulties that we discuss below, that beset even the simplest case of inference about causes of effects, will be hugely magnified when we wish to take additional account of such policy considerations.

5.3 Statistical Evidence

After the above detour, we return to our formulation of the CoE question, in terms of a contrast between R_1, the actually observed response (in this case, $R_1 = 1$) to the treatment actually taken ($E = 1$), and R_0, the (necessarily unknown) counterfactual response, that would have been observed had Ann in fact not taken the aspirin. If "in counterfact" $R_0 = 1$, then Ann's headache would have disappeared even if she had not taken the aspirin, so I must conclude that it was not the aspirin that cured her. Conversely, if $R_0 = 0$ then I can indeed attribute her cure to her having taken the aspirin. In this way, we formulate the CoE causal question in terms of the contrast between the factual outcome R_1 and the counterfactual outcome R_0.

To address the CoE question I thus need to query R_0. Since R_0 has not been observed, it retains a degree of uncertainty, which I could try to express probabilistically. However, not only have I not observed R_0, there is, now, no way I could ever observe it, since once I have observed R_1, R_0 has becomes a counterfactual quantity, predicated on a condition ($E = 0$) that is counter to known facts ($E = 1$). This logical difficulty leads to a degree of unavoidable ambiguity affecting our ability to address the CoE question.

In evaluating my probabilistic uncertainty, I should condition on all I know. My full knowledge about Ann can be expressed as $(E = 1, R_1 = 1, H)$, where H denotes all the background knowledge I have about Ann, and the other variables are likewise individualised to her. With this understanding, we formally define my PROBABILITY OF CAUSATION as the *conditional probability*:

$$\mathrm{PC}_A = \mathrm{P}_A(R_0 = 0 \,|\, H, E = 1, R_1 = 1) \tag{3}$$

where P_A denotes my probability distribution over attributes of Ann.

But how can I go about evaluating PC_A, and what other evidence could be used, and how, to inform this evaluation? In particular, how—if at all—could I make use of EoC probabilities such as (1) and (2) to assist my evaluation of the CoE probability (3)?

5.4 Bounding the Probability of Causation

We note that (3) involves a joint distribution of (R_0, R_1). Since, as a matter of definition, it is never possible to observe both R_0 and R_1 on the same individual, it is problematic to estimate such a joint distribution. We might, however, have a hope of assessing separate

marginal probabilities for R_0 and R_1; and this information can be used to set bounds on PC. Indeed it is straightforward to show (cf. Dawid (2011)):

$$\min\left\{1, \frac{\mathrm{P}_A(R_0 = 0 \,|\, H, E = 1)}{\mathrm{P}_A(R_1 = 1 \,|\, H, E = 1)}\right\} \geq \mathrm{PC}_A \geq \max\left\{0, 1 - \frac{1}{\mathrm{RR}_A}\right\}, \tag{4}$$

where

$$\mathrm{RR}_A := \frac{\mathrm{P}_A(R_1 = 1 \,|\, H, E = 1)}{\mathrm{P}_A(R_0 = 1 \,|\, H, E = 1)}. \tag{5}$$

Readers will recognize (4) as a version of the Bonferroni–Fréchet–Hoeffding bounds (Bonferroni, 1936; Fréchet, 1940; Hoeffding, 1940) that play important rôles in other areas of statistics such as in the study of copulas.

The inequality (4) will yield a non-trivial lower bound so long as $\mathrm{RR}_A > 1$, which we can interpret as saying that there is a positive causal effect of exposure on outcome: cf. the related argument in Robins and Greenland (1989a). Whenever RR_A exceeds 2, we can deduce from (4), without making any further assumptions, that PC_A must exceed 50%. In a civil legal case such as that of Section 2.4, causality might then be concluded "on the balance of probabilities." It is, however, important to note (see Robins and Greenland (1989b)) that, when $\mathrm{RR}_A < 2$, it would not be correct to conclude from this that $\mathrm{PC}_A < 50\%$ (which would lead to the case failing); rather, we can only say that we cannot be sure that the probability of causation exceeds 50%.

The upper bound in (4) is more subtle. It is less than 1 when $\mathrm{P}_A(R_0 = 1 \,|\, H, E = 1) + \mathrm{P}_A(R_1 = 1 \,|\, H, E = 1) > 1$. This happens in general only when both Ann's potential outcomes have a substantial probability of taking value 1. If $\mathrm{P}_A(R_1 = 1 \,|\, H, E = 1) = \mathrm{P}_A(R = 1 \,|\, H, E = 1)$ is only modest in size, e.g. less than 1/2 and $\mathrm{RR}_A > 1$, then the upper bound is 1. If RR_A is large, e.g. $\mathrm{RR}_A > 10$, the upper bound will again be 1 unless $\mathrm{P}_A(R = 1 \,|\, H, E = 1)$ is close to 1. For the remainder of the paper, for simplicity we proceed using an upper bound of 1. Thus we work with the bounds

$$1 \geq \mathrm{PC}_A \geq \max\left\{0, 1 - \frac{1}{\mathrm{RR}_A}\right\} \tag{6}$$

with RR_A given by (5).

5.5 The Risk Ratio

Expression (3) and the denominator of (5) involve a counterfactual consideration: of R_0, Ann's potential response were she not to have taken the aspirin, in the situation that she is known to have taken aspirin ($E = 1$). So it would seem problematic to attempt to identify these quantities from data. However, if my background knowledge H of Ann (on which my distribution P_A is being conditioned) is sufficiently detailed, then, at the point before Ann has decided whether or not to take the aspirin, it might seem appropriate to consider that my uncertainty, conditional on H, about the way her treatment decision E will be made would not further depend on the (so far entirely unobserved) potential responses (R_0, R_1). That is, in this case we might assume

$$(R_0, R_1) \perp\!\!\!\perp_A E \,|\, H \tag{7}$$

where $\perp\!\!\!\perp_A$ denotes conditional independence (Dawid, 1979) in my distribution P_A for Ann's characteristics. When (7) holds we will term the background information H *sufficient*. Then (3) becomes

$$\mathrm{PC}_A = \mathrm{P}_A(R_0 = 0 \,|\, H, R_1 = 1) \tag{8}$$

and in the lower bound in (6) we can replace (5) by

$$\mathrm{RR}_A = \frac{\mathrm{P}_A(R_1 = 1 \,|\, H)}{\mathrm{P}_A(R_0 = 1 \,|\, H)}, \tag{9}$$

my *causal risk ratio* for Ann.[5]

Sufficiency is a kind of "no confounding" requirement on my distribution P_A for Ann (see, for instance, Dawid (2015)). It would fail if, for example, I thought that Ann might take the treatment if she felt really poorly, and not otherwise; but I did not initially have information as to how she felt. Then observing that she took the treatment ($E = 1$) would inform me that she was feeling poorly, so decreasing the probability of a good response (whether actual, R_1, or counterfactual, R_0). Now if I myself am Ann, my H will already include my own knowledge of my perceived state of health, so this argument does not apply, and sufficiency is an acceptable condition.[6] If I am an external observer, however, the sufficiency condition is much more problematic, since I must be able to satisfy myself that my knowledge H of Ann is complete enough to avoid the above possibility of confounding. If I cannot assume sufficiency, I cannot replace the counterfactual denominator of (5) by anything even potentially estimable from data.

Note that the "no confounding" property of sufficiency relates solely to Ann and my knowledge of her. It should not be confused with the superficially similar no confounding property of *exogeneity* described in Section 5.6 below, which refers, not to Ann, but to the process whereby possibly relevant data on other individuals have been gathered.

5.6 Estimating the Risk Ratio

Henceforth we assume sufficiency, which at least gets us started, and we aim to see what further progress can be made, and under what conditions, to get a handle on the bounds on PC_A supplied by RR_A. It is important to be explicit about the assumptions required, which can be very strong and not easy to justify!

It would be valuable if the probabilities featuring in (5) could be related in some way to chances such as (1) and (2) that are estimable from data. Consider first the numerator, the Ann-specific probability $\mathrm{P}_A(R_1 = 1 \,|\, H, E = 1) = \mathrm{P}_A(R = 1 \,|\, H, E = 1)$. It is tempting to replace this by the analogous chance, $\Pr(R = 1 \,|\, H, E = 1)$, which could

[5] We can derive (9)—though not (8)—from the weaker condition that replaces the joint property (7) by the two marginal properties $R_j \perp\!\!\!\perp_A E \,|\, H$, $j = 0, 1$. Since we are only concerned with bounds in this paper, that weaker condition would be adequate for our purposes. However, we find it hard to imagine circumstances where we would be willing to accept the weaker but not the stronger condition, so will continue to use conditions like (7).

[6] This relates to Ramsey's conception of an individual's choice of action as the "ultimate contingency," entirely independent of pre-existing variables given the individual's self-knowledge—see Price (1992).

be estimated from data as for (1), based on the subset of treated trial subjects sharing the same H-value as Ann.[7] This would be justified if we could make the following *bold assumption*:

Condition 1. *Conditional on my knowledge of the pre-treatment characteristics of Ann and the trial subjects, I regard Ann's potential responses as exchangeable with those of the treated subjects having characteristic H.*

Up to this point we have not needed the assumption that H is sufficient. But consider now the denominator of (5). Because of its counterfactual nature, we cannot argue directly as above. However, with sufficiency of H we have $\mathrm{P}_A(R_0 = 1 \,|\, H, E = 1) = \mathrm{P}_A(R_0 = 1 \,|\, H, E = 0) = \mathrm{P}_A(R = 1 \,|\, H, E = 0)$; and we can estimate this from the clinical trial data, e.g. as the estimated chance $\Pr(R = 1 \,|\, H, E = 0)$, if we can assume:

Condition 2. *Conditional on my knowledge of the pre-treatment characteristics of Ann and the trial subjects, I regard Ann's potential responses as exchangeable with those of the untreated subjects having characteristic H.*

Now if both Condition 1 and Condition 2 are to hold, then (by Euclid's first axiom, "Two things that are equal to the same thing are also equal to each other"), the groups of trial subjects with Ann's characteristics H in both arms must be exchangeable with each other. This requires that H be *exogenous*, in the sense that, conditional on H, the potential outcomes (R_0, R_1) have the same distribution among treated and untreated study subjects. This will hold for a suitably randomised study, and also in certain observational studies where the possibility of further confounding factors can be discounted.

Note, however, that we can not take, as H, just *any* exogenous set of variables. The full set of required conditions is:

1. H is exogenous.
2. H is sufficient for Ann's response.
3. Conditional on H, Ann's potential responses are exchangeable with those of the trial subjects.

We will refer to this set of conditions as the *fundamental conditions*.

When we can make good arguments for the acceptability of these fundamental conditions, equation (3) becomes

$$\mathrm{PC}_A = \Pr(R_0 = 0 \,|\, H, R_1 = 1), \tag{10}$$

and, in the lower bound in (4), we can identify RR_A with the population counterpart of (9), the *observational risk ratio*:

[7] Alternatively, the estimate might be constructed from a model for the dependence of the response R on H and $E = 1$, fitted to all the data, and applied with Ann's value of H. We might also be able to reduce to a smaller information set H, if that is all that is relevant for prediction of the responses.

$$\mathrm{ORR} := \frac{\Pr(R = 1 \,|\, H, E = 1)}{\Pr(R = 1 \,|\, H, E = 0)}. \tag{11}$$

Now that we have made clear that the fundamental conditions can be expected to hold only in special circumstances, when they will require detailed justification, we shall henceforth confine ourselves to further consideration of just these special cases. In particular, we shall accept (10), and $\mathrm{RR}_A = \mathrm{ORR}$ as in (11). So we will use the bounds

$$1 \geq \mathrm{PC} \geq \max\left\{0, 1 - \frac{1}{\mathrm{ORR}}\right\} \tag{12}$$

with ORR given by (11). (Here and henceforth, unless the context requires otherwise, we drop the identifier A on PC: these bounds will apply to any individual for whom the fundamental conditions hold.)

5.7 An Alternative Approach

Our probability of causation, PC_A given by (3), is essentially the same as what Pearl (2009, Chapter 9) terms the "Probability of Necessity" (PN). Tian and Pearl (2000) take an alternative approach to supplying bounds for PN, based on data and assumptions different from ours. In particular, they drop our requirement that H be sufficient for Ann's response, requiring instead the availability of two sets of data on individuals exchangeable with Ann: one set in which treatment was (or can be regarded as) randomised, and another in which it arose "naturally" in the same way as for Ann. Because of these differences it is not in general possible to compare their bounds and ours. See Pearl (2015); Dawid et al. (2015) for further discussion of these issues.

5.8 Uncertain Exposure

So far we have supposed we know both the fact of exposure ($E = 1$) and the fact of response ($R = 1$), the only uncertainty being about whether there was a causal link between these two facts. There are other situations where we might observe the response, and wonder whether it was caused by exposure, without knowing with certainty whether or not that exposure had in fact taken place. In such cases we have to multiply the probability of causation PC_A by the probability of exposure, conditional on the known fact of a positive response, yielding a modified probability of causation:

$$\mathrm{PC}^*_A = \mathrm{PC}_A \times \mathrm{P}_A(E = 1 \,|\, H, R = 1). \tag{13}$$

In particular, under the fundamental conditions, combining this with (12) delivers the inequalities

$$\Pr(E = 1 \,|\, H, R = 1) \geq \mathrm{PC}^* \geq \max\left\{0, 1 - \frac{\Pr(E = 0 \,|\, H, R = 1)}{\Pr(E = 0 \,|\, H)}\right\}. \tag{14}$$

6 Statistical Uncertainty

Our discussion so far has treated estimates, such as those in (1) and (2), as if they were the true values of the chances. Even so, we found that we obtain, at best, only partial CoE information, which confines PC or PC* to an interval but does not yield a point value. In real applications our data will not be extensive enough to give us pinpoint estimates of even the bounds featuring in the inequality formulae (12) or (14), and so we have to take additional account of the resulting statistical uncertainty. The result of our inference is thus an *uncertain interval* within which a *probability* (PC or PC*) must lie—thus compounding three different kinds of uncertainty.

Statistical uncertainty, at least, is well studied, and can be expressed and understood in a variety of different ways, as touted and debated by the various competing schools of statistical inference. The generic problem of inference for a quantity (like PC) that, being only partly identified by the data, is subject to interval bounds, has been treated from both a classical perspective (Manski, 2003, 2007; Vansteelandt et al., 2006) and a Bayesian perspective (Greenland, 2005, 2009; Gustafson, 2005, 2009), but these approaches usually involve adding assumptions or data via model expansion and are not directly applicable here.

We here take a Bayesian approach,to derive a joint posterior probability distribution (which, following the helpful terminology of Best et al. (2013), we henceforth term a *credence distribution*) for the estimable unknown chances in the problem.

One possible Bayesian tactic would be to assign a prior distribution to the multivariate parameter, ϕ say, comprising the chances assigned to the four configurations of (R_0, R_1) conditioned on H. Under the fundamental conditions, PC_A is a function of ϕ, given by (10), so that a Bayesian analysis, based on such a prior, would deliver a fully determined posterior distribution for PC_A. However, this is problematic: because R_0 and R_1 are never simultaneously observable, these joint chances cannot be consistently estimated from data, so that this "inference" remains highly sensitive to the specific prior assumptions made, however extensive the data. Alternatively put, the parameter ϕ describing the joint distribution of (R_0, R_1) (given H) is not fully identifiable from data; at best, only the parameter, λ say (a non-invertible function of ϕ), determining the associated marginal distributions of R_0 and R_1, is identifiable. Then λ is a *sufficient parameter* (Barankin, 1960). For extensive data (and a non-dogmatic prior), the posterior distribution of λ will converge to a point mass at its true value, but the posterior conditional distribution of ϕ given λ will be exactly the same in the posterior as in the prior (Kadane, 1974; Dawid, 1979). In particular the marginal posterior distribution of any non-identifiable function of ϕ will be non-degenerate, and highly dependent on the form of the conditional prior for ϕ given λ (Gustafson, 2005, 2009, 2012).

For these reasons, we prefer to assign a joint credence distribution for the (estimable) marginal chances alone: given sufficient data, of sufficiently good quality, these will be well estimated and insensitive to prior assumptions. The price of this increased statistical precision, however, is logical imprecision, since from even perfect knowledge of these chances we can at best derive interval inequalities for PC or PC*. Thus our inference has the form of a *random interval* asserted to contain PC or PC*.

A referee has criticised our unwillingness to work with a distribution for the full parameter ϕ, so as to obtain a well-determined posterior distribution for PC_A. His/her argument is that—quite irrespective of any identifiability considerations—the principles of Bayesian decision theory, and specifically the requirement of Ramsey and de Finetti that one place bets in such a way as to avoid dutch book, demand this. However, the dutch book argument applies only to bets that can (at least in principle) eventually be settled. In our case, the "event" of principal interest, that the response was *caused* by the exposure—involving as it does a comparison between outcomes in two different worlds, in which there either was, or was not, exposure—is undecidable, even in principle. One might even take the extreme philosophical position that there is no "actual" exact value of PC_A, and that it simply has no more precise existence beyond what is expressed by its interval bounds, which relate to decidable events.

Nevertheless we must admit that, in the legal context, the trier of fact may be required to make a decision about causality. Taking account of the adversarial nature of the legal process, we would argue that, on the basis of the quantitative evidence alone, the trier of fact should only infer causality ("on the balance of probabilities") when the lower bound on PC or PC^* exceeds 50% (with high probability). Conversely, lack of causation is established only when the upper bound falls below 50%. In intermediate cases, where the relevant interval straddles 50%, no meaningful inference based solely on the specific quantitative evidence is possible. There may, however, be other relevant evidence that could tip the balance one way or the other.

6.1 Group or Individual Inference?

In the above approach, we considered the probabilities featuring in the inequalities (12) and (14) as "objective chances," which we might interpret as limiting relative frequencies computed in an appropriate groups of exchangeable individuals. We focused on the posterior joint credence distribution of these chances, given the available statistical data D—thus giving rise to a random uncertainty interval for the probability of causation, itself regarded as an objective chance. We refer to this as the "group-focused" approach.

Another approach to using data to inform the inference about PC or PC^* is to regard these concepts, and the probabilities featuring in the bounds for them, themselves as credences, quantifying numerically the relevant uncertainty about attributes of the specific individual, Ann, on whom we are focusing. This is the "individual-focused" approach. For an interchange on these issues and "group-to-individual" inference, see the discussion by Dawid and the authors' rejoinder in Best et al. (2013).

In the individual-focused formulation, the term $\Pr(E = 0 \,|\, H, R = 1)$ in (14), for example, would be replaced by $\mathrm{P}_A(E_A = 0 \,|\, H_A, R_A = 1, D)$, where the suffix A refers to attributes of Ann. We now obtain a non-random uncertainty interval for PC_A (or PC^*_A)—but one that is computed in the light of the available evidence, and would be likely to change were further data to become available.

To continue with this example, let ψ denote the chance $\Pr(E = 0 \,|\, H, R = 1)$. If the individuals in D can be regarded as exchangeable with Ann, and we interpret ψ as a limiting relative frequency in this exchangeable setting, we will have:

$$\begin{aligned} \mathrm{P}_A(E_A = 0 \,|\, H_A, R_A = 1, D) &= \mathrm{E}\left\{\mathrm{P}_A(E_A = 0 \,|\, H_A, R_A = 1, \psi) \,|\, H_A, R_A = 1, D\right\} \\ &= \mathrm{E}\left(\psi \,|\, H_A, R_A = 1, D\right). \end{aligned} \tag{15}$$

Often, given the data D, the further conditioning in (15) on the Ann-specific information $(H_A, R_A = 1)$ will have negligible effect—in which case the desired Ann-specific probability $\mathrm{P}_A(E_A = 0 \,|\, H_A, R_A = 1, D)$ can be approximated by the posterior expectation (i.e. conditioned on D alone) of the conditional chance $\psi = \Pr(E = 0 \,|\, H, R = 1)$. A similar argument applies to any other required credences in the problem.

6.2 Additional Issues

All our above analysis is predicated on the causal relevance of the epidemiological data, assuming that we can use the study to obtain a sound estimate of the causal risk ratio RR_A that features in (5). For example, in a simple fully randomised study we could use ORR, as given by (11), as a proxy for RR. But such studies are the exception in epidemiology, so that the issues in real world settings where interest is focused on the causes of effects are typically much more complex. Thus in the benfluorex example of Section 2, using the frequencies in Table 1 for this purpose, by plugging them into the formula for ORR and interpreting this as RR, would be totally misleading, even if we attempted to account for statistical uncertainty as described above. Indeed, as we noted in Section 2, there are additional problems in this case: because the study of Frachon et al. (2010) was retrospective, and the frequencies in Table 1 could not be used to estimate ORR, even in the absence of confounding. And this problem remains when, admitting the likely existence of confounding, we conduct a more sophisticated analysis—such as the multiple logistic regression that produced the adjusted odds ratio—to try and account for it. Even when this ploy can be regarded as successful, still the best we can ever do with retrospective data is to estimate the causal odds ratio—which will approximate the desired causal risk ratio, as required for setting the lower bound on PC, only when the outcome is rare.

The judge in the hypothetical case we pose should therefore be doubly wary of the relevance of the epidemiological evidence when trying to assess whether the drug caused the plaintiff's heart disease.

There are even more complex situations where the data are retrospective and where there are multiple outcomes of interest and multiple time points for their assessment. A notable example comes from the continuing effort in the United States to examine the long-term health effects of exposure to Agent Orange among US Vietnam veterans. From 1962 to 1971, the US military sprayed herbicides over Vietnam. In 1991 the US Congress passed the Agent Orange Act, requiring a comprehensive evaluation of scientific and medical information regarding the health effects of exposure to Agent Orange and other herbicides used in Vietnam: Institute of Medicine (2011) is the eighth biennial update implementing this Congressional mandate. The report examines epidemiological studies of the health status of veterans considering a multiplicity of deleterious effects, e.g. different forms of cancer and early-onset peripheral neuropathy, and with limited information on exposure, both at the aggregate and individual level. A standard tool in the studies incorporated into this regularly-updated assessment is the use of adjusted

odds-ratios from retrospective logistic regression analyses. Identification of a substantial RR triggers compensation to veterans for health and disability outcomes associated with putative exposure.

7 Case Study

We illustrate our analysis with an example taken from Best et al. (2013). The motivating real life case was the diagnosis of abuse in an infant child, c, presenting with an acute life-threatening event ("ALTE"). So now we take exposure, $E = 1$, to denote abuse, and response, $R = 1$ to denote ALTE.

7.1 Three Tasks

We can distinguish three tasks that we might wish to address probabilistically concerning the relationship between exposure and response in this individual case; these are quite distinct and should not be confused—although there are of course relationships (far from trivial) between them.

Forecasting If the child is abused, what is the probability the child will suffer ALTE—$\mathrm{P}_c(\mathrm{ALTE} \,|\, \mathrm{abuse})$?

Backcasting If the child suffers ALTE, what is the probability the child was abused?—$\mathrm{P}_c(\mathrm{abuse} \,|\, \mathrm{ALTE})$?

Attribution If the child suffers ALTE, what is the probability this was caused by abuse?—PC^*_c, as in (13)?

In the above, we have used P_c to indicate my probabilities for this child (implicitly conditioned on the available data, and on the specific background information H I have about the child). Even so, we have a choice between taking a group-focused approach, in which P_c is interpreted as an uncertain chance, relevant to a group of individuals of which this child is one; or an individual-focused approach, with P_c denoting my credence about this specific child. We start by taking the group-focused approach: the individual-focused approach will be considered in Section 8.3.

7.2 Attribution Analysis

Best et al. (2013) focused on the backcasting task: of assessing whether or not abuse has in fact taken place, based on the data on the individual case and on relevant statistical studies. Their analysis directly addressed the main substantive concern, since it was the occurrence of abuse—whether or not it in fact caused the observed signs—that was at issue. They did not need to enquire whether or not the observed signs were *caused* by abuse. That attribution question, however, will be our focus here. We note that, since the very fact of abuse is itself uncertain, we also need to consider the backcasting issue. This is done by taking, as the relevant probabilistic target of our inference, the modified probability of causation PC^*, as given by (13).

We have described in Section 5.3 the many very strong assumptions that have to be made in order to justify using data to estimate even the weak interval bounds of (14) for PC*. In the present example, the data used by Best et al. (2013) were gleaned from a search for relevant published studies. Those identified were of varying design and quality, and the data extracted from them can in no sense be regarded as supporting genuine causal inferences—indeed, it is not easy to find real examples where the conditions supporting causal inference of this type could be regarded as satisfied. Purely for illustration we shall proceed as if they are, so that we can use the inequalities of (14). As a further—admittedly highly unrealistic—simplifying assumption, we take the sufficient information H to be trivial. All these imperfections in the data, and in our understanding of the context, mean that our analysis must not be taken as delivering a credible conclusion in this particular application; however, we hope that, by following it through in detail, we may help to clarify the points to which attention should given when analysing any similar problem.

Using a Gibbs sampler implemented in the WinBUGS © software (Lunn et al., 2012), Best et al. (2013) find the posterior credence distributions for various conditional chances, based on the data.[8] In particular, they obtain the posterior credence distribution for the conditional chance $\Pr(E = 1 \,|\, R = 1)$, and thus for $\Pr(E = 0 \,|\, R = 1) = 1 - \Pr(E = 1 \,|\, R = 1)$, as needed on both sides of (14).

For our purposes, however, we need more: the lower bound for PC* in (14) also involves the marginal prior chance $\Pr(E = 0)$—or, essentially equivalently, $\theta := \Pr(E = 1)$, the chance of abuse having taken place (in this individual case), before the evidence of ALTE is taken into consideration. And there is no available statistical evidence relevant to this quantity.

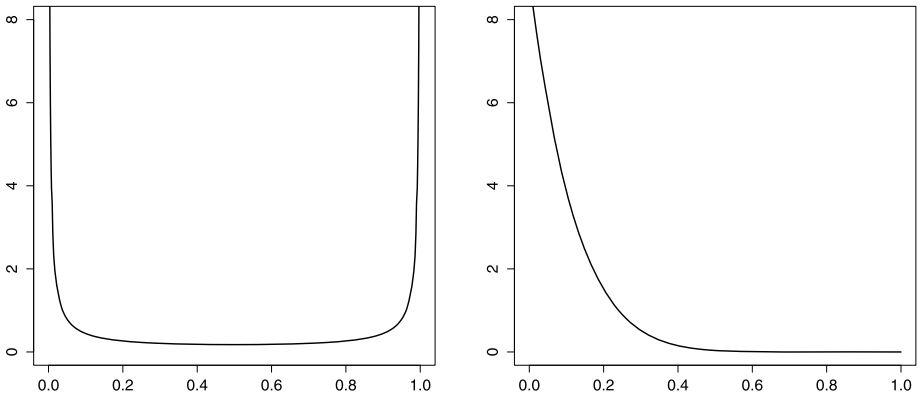
We therefore proceed by introducing our own prior credence distribution for θ, and treating this chance as independent of all the others in the problem. We can expect considerable sensitivity to the specific choice made. To begin to explore this, we try two different prior credence distributions for θ, both beta distributions for simplicity and tractability:

Prior 1: $\theta \sim \beta(0.1, 0.1)$.
This has mean 0.5 and standard deviation 0.46. It can be regarded as representing very substantial prior uncertainty about θ: its mean 0.5 is the unconditional probability assigned to abuse.

Prior 2: $\theta \sim \beta(1, 9)$.
This has mean 0.1 and standard deviation 0.09. While still admitting uncertainty, it attempts to take into account the prior unlikelihood of abuse: its mean 0.1 is the unconditional probability assigned to this event.

Density functions of these two prior credence distributions are displayed in Figure 1.

[8] Best et al. (2013) conduct several alternative analyses, with some of the less reliable data values being either included or excluded. Our own analysis is based on the predictive model and data in the combined WinBUGS code of Appendices B and D of their paper, as for their own Table 4. This analysis targets a case-specific chance, having greater relevance, but also more uncertainty, than the overall population-based chance.



(a) Prior 1: $\theta \sim \beta(0.1, 0.1)$ (b) Prior 2: $\theta \sim \beta(1, 9)$

Figure 1: Two prior credence distributions for θ, the prior chance that abuse has taken place.

We note that the lower bound in (14) is 0 if and only if $\mathrm{RR}_A \leq 1$, which event is independent of θ. Consequently, the posterior probability that the lower bound is 0 is unaffected by the assumptions made about the prior distribution for θ—as similarly is the conditional posterior distribution of the upper bound, given that the lower bound is 0.

8 Data Analysis

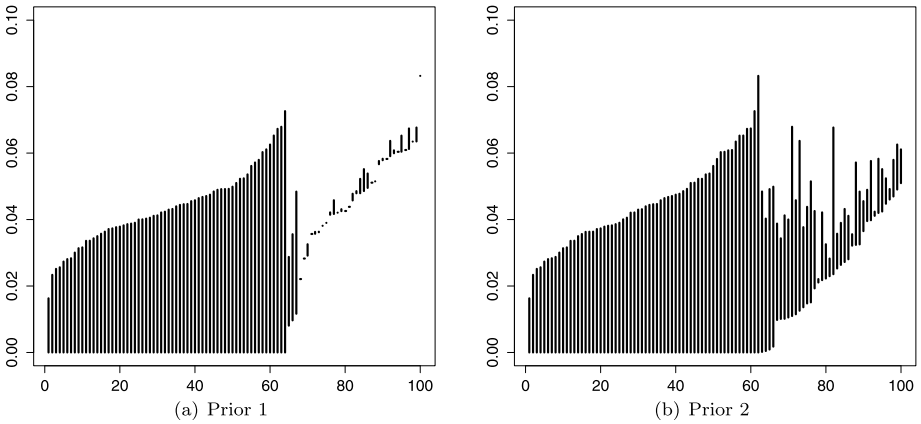
We have conducted our own analysis of the data, based on the WinBUGS code of Best et al. (2013) elaborated so as to incorporate θ. After a burn-in phase of 500000 iterations, to get rid of autocorrelation we have based the estimates on thinned samples taking every 10th element of the chain. Then we have considered a chain of length 50000.[9]

8.1 Bivariate Distribution

A complete inference would describe the posterior credence distribution of the uncertainty interval (14) for PC^*, whose end-points are functions of random chances, and hence themselves have a bivariate distribution.

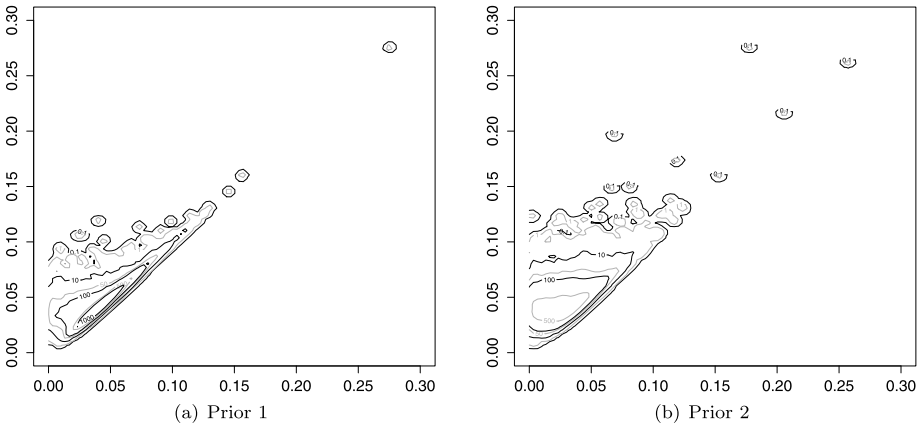
Note that, whenever the inequality $\Pr(E = 1 \,|\, R = 1) \leq \Pr(E = 1)$ between chances holds, which corresponds to negative association between exposure and outcome and will happen with positive probability in the posterior credence distribution, the lower bound of the uncertainty interval is 0 and is thus entirely uninformative. Thus the posterior credence distribution is a mixture of a continuous bivariate distribution, and

[9] This is almost certainly overkill, and should not be taken as a recommendation.



(a) Prior 1 (b) Prior 2

Figure 2: For each of the priors of Figure 1, 100 uncertainty intervals, randomly sampled from the bivariate posterior distribution of the lower and upper bounds, are displayed. The intervals are ordered in increasing value of the lower bound.



(a) Prior 1 (b) Prior 2

Figure 3: For each of the priors of Figure 1, a contour plot of the joint posterior distribution of the lower and upper bounds of the random uncertainty interval, conditional on the lower bound being positive.

(with positive probability) a distribution for the upper bound alone. The probability that the lower bound of the uncertainty interval is 0 (which is independent of the prior distribution used) is estimated as 0.65. Figure 2 displays, for the two different priors, samples from the bivariate posterior credence distribution (ordered by lower bound). In the plots 100 uncertainty intervals are reported, obtained by selecting one iteration of the chain every 500.

In Figure 3 bivariate contour plots of the end-points of the random uncertainty interval are shown for Priors 1 and 2, excluding those cases where the lower bound is

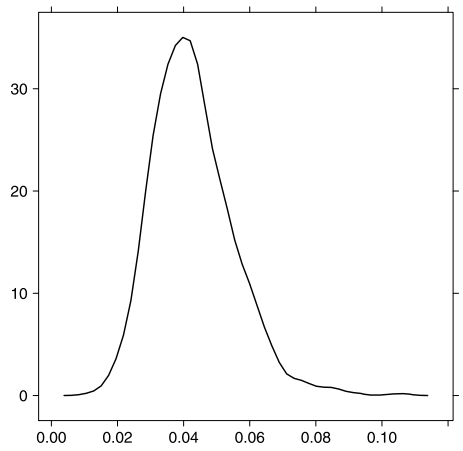


Figure 4: The posterior density of the upper bound of the random uncertainty interval, conditional on the lower bound being 0. This distribution is independent of the chosen prior for the parameter θ.

equal to zero. The full joint distribution is completed by specifying the distribution of the upper bound for these. This distribution, which is independent of the assumed prior for θ, is shown in Figure 4, which can also be interpreted as displaying the conditional distribution of the length of the uncertainty interval for PC*, given that its lower bound is 0.

8.2 Univariate Summaries

Useful univariate summaries of the overall bivariate inference are the marginal posterior credence distributions of the upper and lower bounds, and of the length of the uncertainty interval.

Upper Bound

The upper bound $\Pr(E = 1 \mid R = 1)$ in (14) is the chance of abuse given the case evidence of ALTE, as already considered by Best et al. (2013). Its posterior credence distribution (which is unaffected by the choice of prior for θ) is summarised in the second row of Table 4 of Best et al. (2013). We compute the posterior mean and standard deviation for this upper bound to be 0.043 and 0.013, respectively. Its posterior density is shown in Figure 5.

Lower Bound

The lower bound on PC* in (14), $\max\{0, 1 - \Pr(E = 0 \mid R = 1)/\Pr(E = 0)\}$, depends also on $\theta = \Pr(E = 1)$, and its posterior credence distribution could be sensitive to

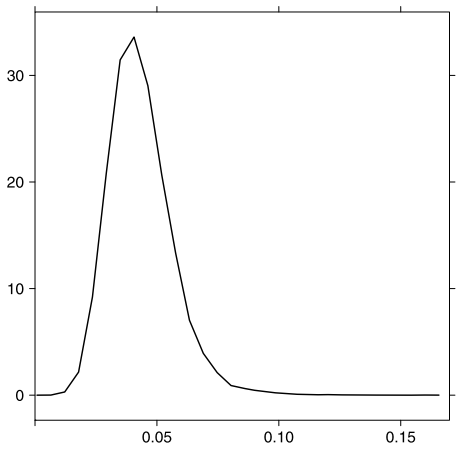


Figure 5: The posterior credence density of the upper bound for PC*. This distribution is independent of the chosen prior for the parameter θ.

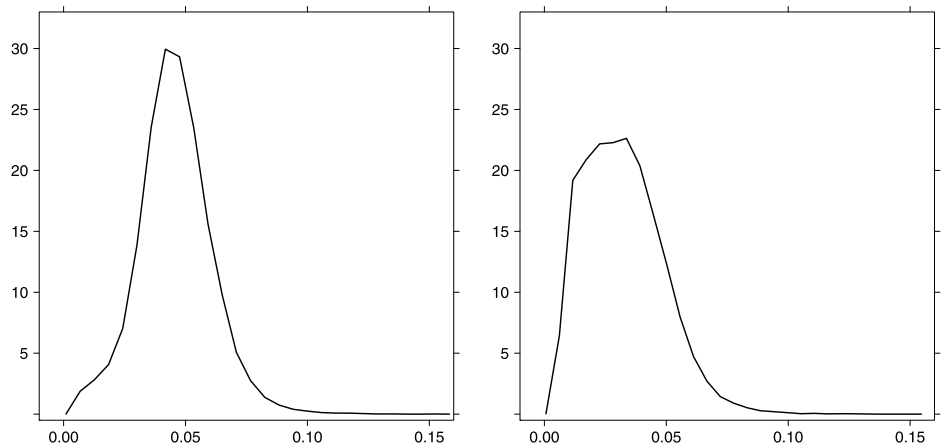
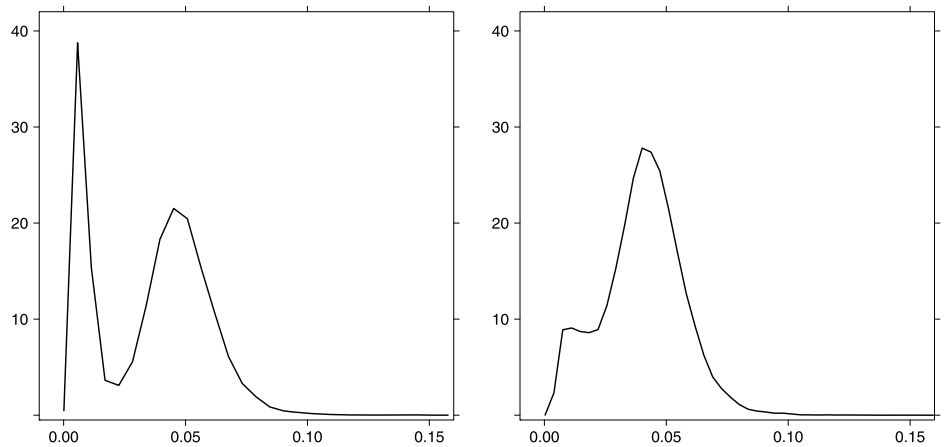


Figure 6: For each of the priors of Figure 1, the posterior credence density of the lower bound for PC*, conditional on this being greater than 0.

the prior credence distribution chosen for θ. We have already noted that the posterior credence probability that the lower bound is 0 is 0.65, independent of the prior for θ. Figure 6 displays the posterior densities for the lower bound, conditional on its being strictly positive, for Prior 1 and Prior 2; the means are 0.039 and 0.025, and the standard deviations are 0.015 and 0.016, respectively. We see that the effects of the differences between the priors are relatively minor.



(a) Prior 1 (b) Prior 2

Figure 7: For each of the priors of Figure 1, the posterior credence density of the length of the uncertainty interval for PC*.

Length of Interval

Another useful summary of the full inference is the posterior credence distribution of the length of the interval between the lower and upper bounds on PC*, as displayed in Figure 7.

The posterior mean and standard deviation based on Prior 1 are, respectively, 0.028 and 0.022, while for Prior 2 these quantities are 0.035 and 0.016. We see high sensitivity to the prior assumptions. This is particularly apparent when we exclude data with lower bound equal to zero (see Figure 8).

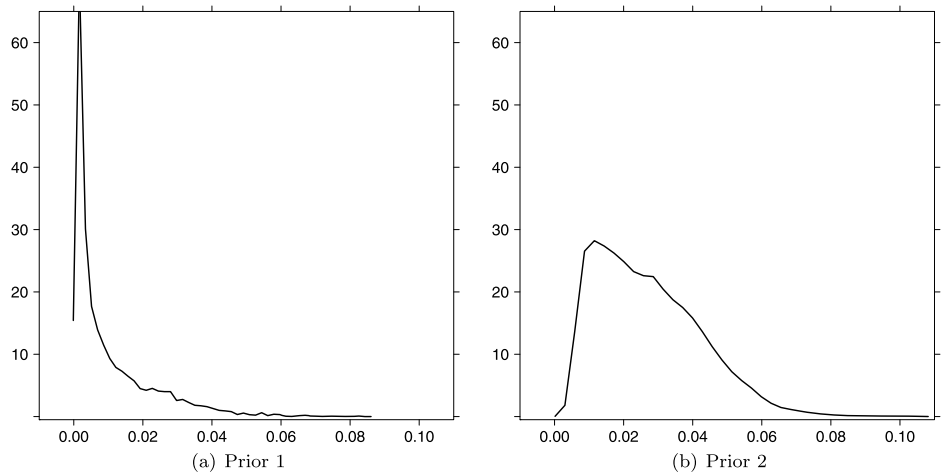
For cases with lower bound equal to 0, the interval length is identical with the upper bound, as displayed in Figure 4, and is independent of the prior distribution for θ. These features are also visible in Figure 2.

Coverage Probability

Finally, for any probability value p, we can compute the posterior credence that this is included in the random interval (14)—and thus is at least a candidate as a value for PC*. We graph this coverage measure, as a function of p, in Figure 9 for both priors.

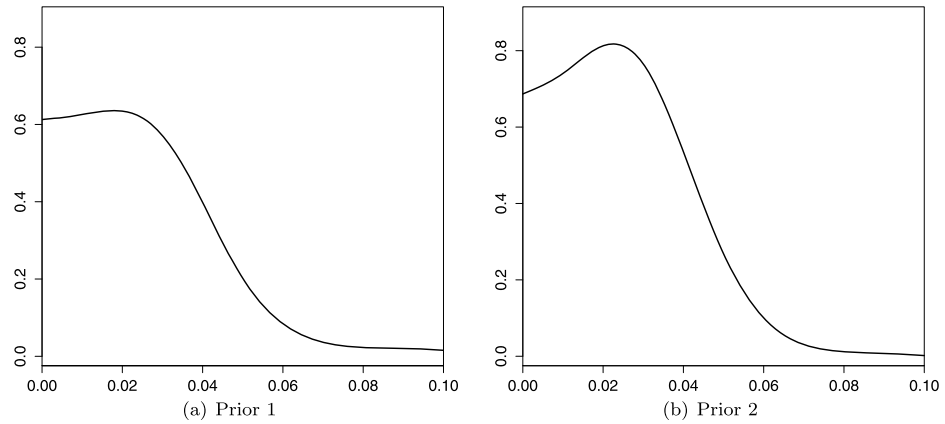
8.3 Individual-Focused Inference

The individual-focused inference is much simpler in form: according to the analysis in Section 6.1 (and assuming the approximation mentioned there is valid), we simply replace the chances featuring in the bounds of (14) by their posterior expectations. (Recall that we are taking H to be trivial, so it can be omitted from the notation).



(a) Prior 1 (b) Prior 2

Figure 8: For each of the priors of Figure 1, the posterior credence density of the length of the uncertainty interval for PC*, conditional on the lower bound being greater than 0.



(a) Prior 1 (b) Prior 2

Figure 9: For each of the priors of Figure 1, the posterior credence probability that the random uncertainty interval covers any specific value.

The posterior expectation of the upper bound, $\Pr(E = 1 \,|\, R = 1)$, is 0.043, independent of the assumed prior distribution for θ.

As for the lower bound, the posterior expectation of $\Pr(E = 0 \,|\, R = 1)$ is $1 - 0.043 = 0.957$. Also, $\Pr(E = 0) = 1 - \theta$, and since we have no data relevant to θ the posterior expectation of this quantity is the same as its prior expectation, namely 0.5 for Prior 1, or 0.9 for Prior 2. It is clear that there could be high sensitivity to the prior distribution assessed for θ. However, in this case the lower bound is 0 for both

priors. Hence our individual-focused uncertainty interval for PC* is $(0, 0.043)$ in both cases.

9 Conclusions

We have seen that statistical inference about "causes of effects" is particularly problematic from many points of view, and difficult to justify even in ideal circumstances.

First, in order merely to formalise the question, we need to carefully specify, separately, both who is making the inference (in § 5.3 we called that person"I") and who (there called "Ann") the inference relates to. Next, we need to be satisfied that my information H about Ann is *sufficient*, in the sense of there being no confounding that could make Ann's treatment choice informative (for me) about her potential outcome variables. When all these conditions are satisfied we can begin to try and learn from relevant data about the two versions, PC and PC*, of the probability of causation. For that purpose we should have good experimental data from which we can get good estimates of the distribution of the outcome, conditional on exposure E and H. And even with such ideal estimated probabilities, the resulting inferences are complex, compounding as they do three different kinds of uncertainty: interval bounds, for a probability, that are themselves random. We have made a start at exploring ways of understanding, describing and displaying such triple uncertainty (in an example that admittedly falls far short of the ideal situation), but much remains to be done.

References

Altham, P. M. E. (1970). "The Measurement of Association of Rows and Columns for an $r \times s$ Contingency Table." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 32: 63–73. http://www.jstor.org/stable/2984403 728

Anon (17 November 2010). "500 Deaths Linked To Diabetes Drug Mediator, France." *Medical News Today*. http://www.medicalnewstoday.com/articles/208343.php 727

Barankin, E. W. (1960). "Sufficient Parameters: Solution of the Minimal Dimensionality Problem." *Annals of the Institute of Statistical Mathematics*, 12: 91–118. MR0126308. doi: http://dx.doi.org/10.1007/BF01733119. 737

Best, N., Ashby, D., Dunstan, F., Foreman, D., and McIntosh, N. (2013). "A Bayesian Approach to Complex Clinical Diagnoses: A Case-Study in Child Abuse (with Discussion)." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176: 53–96. MR3042177. doi: http://dx.doi.org/10.1111/j.1467-985X.2012.01050.x. 737, 738, 740, 741, 742, 744

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press. Reprinted by Springer-Verlag, New York (2007). MR2344876. 728

Bonferroni, C. E. (1936). "Teoria Statistica delle Classi e Calcolo delle Probabilità." *Publicazioni del R. Instituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8: 1–62. 733

Dawid, A. P. (1979). "Conditional Independence in Statistical Theory (with Discussion)." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 41: 1–31. MR0535541. 734, 737

— (2011). "The Role of Scientific and Statistical Evidence in Assessing Causality." In: Goldberg, R. (ed.), *Perspectives on Causation*, 133–147. Oxford: Hart Publishing. 729, 730, 733

— (2015). "Statistical Causality from a Decision-Theoretic Perspective." *Annual Review of Statistics and Its Application*, 2: 273–303. doi: http://dx.doi.org/10.1146/annurev-statistics-010814-020105. 734

Dawid, A. P., Faigman, D. L., and Fienberg, S. E. (2014). "Fitting Science Into Legal Contexts: Assessing Effects of Causes or Causes of Effects? (with Discussion)." *Sociological Methods & Research*, 43: 359–390. MR3248013. doi: http://dx.doi.org/10.1177/0049124113515188. 728, 729

— (2015). "On the Causes of Effects: Response to Pearl." *Sociological Methods & Research*, 44: 165–174. MR3313823. doi: http://dx.doi.org/10.1177/0049124114562613. 736

Farewell, V. (1979). "Some Results on the Estimation of Logistic Models Based on Retrospective Data." *Biometrika*, 66: 27–32. MR0529144. doi: http://dx.doi.org/10.1093/biomet/66.1.27. 728

Fienberg, S. E. and Zellner, A. (eds.) (1974). *Studies in Bayesian Econometrics and Statistics. Essays In Honor of Leonard J. Savage*. North-Holland. 750

Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd. 730

Frachon, I., Étienne, Y., Jobic, Y., Le Gal, G., Humbert, M., and Leroyer, C. (2010). "Benfluorex and Unexplained Valvular Heart Disease: A Case-Control Study." *PLOS One*, 5(4): e10128. doi: http://dx.doi.org/10.1371/journal.pone.0010128. 726, 727, 728, 729, 739

Frachon, I. N. (2010). *Médiator 150mg: Combien de Morts?*. Brest: editions-dialogues.fr. http://www.editions-dialogues.fr/livre/mediator-150-mg/ 726

Fréchet, M. (1940). *Les Probabilités Associées à un Système d'Évènements Compatibles et Dépendants*, volume 1. Paris: Hermann & Cie. 733

Gelman, A. and Imbens, G. (2013). "Why ask Why? Forward Causal Inference and Reverse Causal Questions." Working Paper 19614, National Bureau of Economic Research, Cambridge, MA. http://www.stat.columbia.edu/~gelman/research/unpublished/reversecausal_13oct05.pdf 729

Green, M. D., Freedman, M. D., and Gordis, L. (2011). "Reference Guide on Epidemiology." In: *Reference Manual on Scientific Evidence: Third Edition*, 549–632. Washington, D.C.: The National Academies Press. 725

Greenland, S. (2005). "Multiple-Bias Modelling for Analysis of Observational Data (with Discussion)." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168: 267–306. MR2119402. doi: http://dx.doi.org/10.1111/j.1467-985X.2004.00349.x. 737

— (2009). "Relaxation Penalties and Priors for Plausible Modeling of Nonidentified Bias Sources." *Statistical Science*, 24: 195–210. MR2655849. doi: http://dx.doi.org/10.1214/09-STS291. 737

Gustafson, P. (2005). "On Model Expansion, Model Contraction, Identifiability and Prior Information: Two Illustrative Scenarios Involving Mismeasured Variables." *Statistical Science*, 20: 111–140. MR2183445. doi: http://dx.doi.org/10.1214/088342305000000098. 737

— (2009). "What Are the Limits of Posterior Distributions Arising From Nonidentified Models, and Why Should We Care?" *Journal of the American Statistical Association*, 104: 1682–1695. MR2750585. doi: http://dx.doi.org/10.1198/jasa.2009.tm08603. 737

— (2012). "On the Behaviour of Bayesian Credible Intervals in Partially Identified Models." *Electronic Journal of Statistics*, 6: 2107–2124. MR3020258. doi: http://dx.doi.org/10.1214/12-EJS741. 737

Hill, A. B. (1951). "The Clinical Trial." *British Medical Bulletin*, 7: 278–82. 730

Hill, C. (2011). "Mortalité Attribuable au Benfluorex (Mediator®)." *La Presse Médicale*, 40: 462–469. doi: http://dx.doi.org/10.1016/j.lpm.2011.03.001. 726

Hoeffding, W. (1940). "Scale-Invariant Correlation Theory." *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5: 181–233. 733

Holland, P. W. (1986). "Statistics and Causal Inference (with Discussion)." *Journal of the American Statistical Association*, 81: 945–970. MR0867618. doi: http://dx.doi.org/10.2307/2289064. 729

Holland, P. W. and Rubin, D. B. (1988). "Causal Inference in Retrospective Studies." *Evaluation Review*, 12: 203–231. doi: http://dx.doi.org/10.1177/0193841X8801200301. 727, 730

Institute of Medicine (2011). *Committee to Review the Health Effects in Vietnam Veterans of Exposure to Herbicides (Eighth Biennial Update): Veterans and Agent Orange: Update 2010*. Washington, DC: The National Academies Press. 739

Kadane, J. B. (1974). "The Rôle of Identification in Bayesian Theory." In: Fienberg and Zellner (1974), 175–91. MR0483124. 737

Laplace, P. S. (1986). "Memoir on the Probability of the Causes of Events." *Statistical Science*, 1: 364–378. MR0858515. 728

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Texts in Statistical Science. Chapman & Hall/CRC. 726, 741

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. New York: Springer-Verlag. MR2151380. 737

— (2007). *Identification for Prediction and Decision*. Cambridge, MA: Harvard University Press. 737

Mill, J. S. (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and Methods of Scientific Investigation*. London: John W. Harper. 729

Nicot, P., Frachon, I., and Hill, C. (2011). "À Propos de l'Expertise des Dossiers Benfluorex (Mediator® et Génériques)." *La Presse Médicale*, 42(4): 411–418. doi: http://dx.doi.org/10.1016/j.lpm.2013.02.310. 728

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press, second edition. MR2548166. doi: http://dx.doi.org/10.1017/CBO9780511803161. 730, 736

— (2015). "Causes of Effects and Effects of Causes." *Sociological Methods & Research*, 44: 149–164. MR3313822. doi: http://dx.doi.org/10.1177/0049124114562614. 736

Price, H. (1992). "The Direction of Causation: Ramsey's Ultimate Contingency." In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 2: Symposia and Invited Papers, 253–267. The University of Chicago Press on behalf of the Philosophy of Science Association. http://www.jstor.org/stable/192840 734

Robins, J. and Greenland, S. (1989a). "Estimability and Estimation of Excess and Etiological Fractions." *Statistics in Medicine*, 8: 845–859. 733

— (1989b). "The Probability of Causation Under a Stochastic Model for Individual Risk." *Biometrics*, 45: 1125–1138. MR1040629. doi: http://dx.doi.org/10.2307/2531765. 733

Rubin, D. B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66: 688–701. 730, 731

Tian, J. and Pearl, J. (2000). "Probabilities of Causation: Bounds and Identification." *Annals of Mathematics and Artificial Intelligence*, 28: 287–313. MR1797625. doi: http://dx.doi.org/10.1023/A:1018912507879. 736

Vansteelandt, S., Goetghebeur, E., Kenward, M. G., and Molenberghs, G. (2006). "Ignorance and Uncertainty Regions as Inferential Tools in a Sensitivity Analysis." *Statistica Sinica*, 16: 953–979. MR2281311. 737

Weill, A., Pata, M., Tuppin, P., Fagot, J. P., Neumann, A., Simon, D., Ricordeau, P., Montastruc, J. L., and Allemond, H. (2010). "Benfluorex and Valvular Heart Disease: A Cohort Study of a Million People with Diabetes Mellitus." *Pharmacoepidemiology and Drug Safety*, 19: 1256–1262. doi: http://dx.doi.org/10.1002/pds.2044. 726, 727

Acknowledgments

We thank Catherine Laurent for valuable background about Irène Frachon and benfluorex. We thank the referees, Associate Editor and Editor for their comments on earlier versions of this paper. The work of Philip Dawid was partially supported by an Emeritus Fellowship from The Leverhulme Trust. The work of Monica Musio was partially supported by the grant no. F71J12001120002 from the Regione Sardegna (Progetto di Ricerca Fondamentale o di Base L.R. 7/2007 annualità 2012). Stephen Fienberg was supported in part by the Singapore National Research Foundation (NRF) under its International Research Centre @ Singapore Funding Initiative and the Interactive Digital Media Programme Office (IDMPO) to the Living Analytics Research Centre (LARC), and by the U.S. National Institute of Standards and Technology through its funding of the Center for Statistics and Applications in Forensic Evidence.