Interpreting Probability in Causal Models for Cancer

Federica Russo and Jon Williamson

ABSTRACT. How should probabilities be interpreted in causal models in the social and health sciences? In this paper we take a step towards answering this question by investigating the case of cancer in epidemiology and arguing that the objective Bayesian interpretation is most appropriate in this domain.

After introducing the problem in §1 and giving an overview of causal analysis in the social and health sciences in §2, in §3 we present the cancer case study in some detail. In §4 we discuss the importance of correctly interpreting probability. Then, in §5, we put forward some desiderata that an interpretation of probability ought to satisfy; two Bayesian interpretations of probability come out well according to these desiderata. In §6 we go further by showing how the full-blown objectivity of objective Bayesianism is needed for the practice of cancer treatment. Finally we discuss the ramifications of this conclusion for the social and health sciences in §7.

1 Introduction

Whilst it might seem uncontroversial that the health sciences search for causes—that is, for causes of disease and for effective treatments—the causal perspective is less obvious in the social sciences, perhaps because it is apparently harder to glean general laws in the social sciences than in other sciences. Thus the search for causes in the social sciences is often perceived to be a vain enterprise and it is often thought that social studies merely describe the phenomena.

On the other hand an explicit causal perspective can already be found in pioneering works of Adolphe Quetelet and Emile Durkheim in demography and sociology respectively, and the social sciences have taken a significant step in *quantitative causal analysis* by following Sewall Wright's path analysis, which was first applied in population genetics. Subsequent developments of path analysis—e.g. structural models, covariance structure models and multilevel analysis—have the merit of making the concept of

cause operational by introducing causal relations into the framework of statistical modelling. However, these developments in causal modelling leave a number of conceptual issues unanswered: for instance the question of how probability should be interpreted in probabilistic causal models.

In the philosophy of probability many interpretations have been proposed and crucial objections raised. For instance, it has been argued that the frequency interpretation does not make sense in the single-case, that subjectivist accounts lead to arbitrariness in probability assignments, and that logical interpretations, though suited to gambling situations, are of scarce applicability in science.

In this paper we raise the problem of the interpretation of probability within a specific context: causal models in cancer epidemiology.[1] This is motivated by the thought that competing interpretations are not right or wrong, but that they are better or worse suited to particular contexts and the demands we make of them. To this end, we first introduce causal analysis in the social and health sciences and then present the case of cancer epidemiology in some detail. We pay particular attention to explaining different possible meanings of probabilistic statements in this context and the importance of choosing one interpretation of probability over another. We then argue that any satisfactory interpretation of probability should satisfy five desiderata; this narrows down the choice to the frequency interpretation twinned with an empirically-based subjective interpretation or with an objective Bayesian interpretation. We go on to argue that the probabilities in causal models in cancer epidemiology should be given a frequency-cum-objective-Bayesian interpretation; the main reason for this choice is the need to cope with two different types of probabilistic inference, generic and single-case.

2 Causality in the Social and Health Sciences

Different social sciences study society from different perspectives. Sociology studies the structure and development of human society, demography attends to the vital statistics of populations, economics studies the management of goods and services, epidemiology studies the distribution of disease in human populations and the factors determining that distribution, etc. In spite of these differences, the social and health sciences share a common objective: to understand, predict and intervene on society. Knowledge of

[1] We are fairly liberal as to which models count as *causal*. Arguably, a model is causal if its relationships are interpreted causally or put to causal use. Thus associational and regression models, which tend not be explicitly causal, would count as causal for us if the relationships in the model are interpreted causally (e.g., if an association between smoking and cancer is interpreted as supporting the claim that smoking causes cancer) or used as a basis for intervention (e.g., by banning tobacco advertising).

causes is required to achieve this common goal; such knowledge provides an explanation of social phenomena as well as of individual behaviour.

Whilst social scientists have been looking for social causes, health scientists traditionally focus on the biological causes of diseases. For instance, in *Le Suicide* Emile Durkheim studies suicide as a social phenomenon and consequently looks for the social causes of the suicide rate in the population. Modern medicine, since the work of Claude Bernard (1813-1878), tries to identify physiological mechanisms active in living beings. The Henle-Koch Postulates, formulated by Robert Koch in 1882, provided criteria for judgements about the presence of micro-organisms as causes of disease (Koch, 1882). For instance, the current theory of carcinogenesis involves a particular molecular mechanism. Under normal circumstances the growth of cells is controlled accurately by inherited mechanisms and stimulated or inhibited as required. Cancer occurs when genetic alterations (mutations) disturb the normal regulation of a cell and cause it to erroneously multiply—with potentially fatal consequences. Mutations like this trigger cell growth either directly or indirectly by disrupting the mechanisms responsible for limiting cell division. On the other hand, the social sciences try to understand the variety of ways in which the population is exposed to carcinogenic substances, according to levels of education, economic status, type of occupation, etc.

The association between diseases and various socio-economic factors is the object of research in several disciplines. In particular, epidemiology typically tries to single out individual genetic factors, biological factors, and environmental risk factors. Studies have recently focused on the role of neighbourhood environment (see Pickett and Pearl (2001) for a literature review). Neighbourhood environment affects health through the availability and accessibility of health services, infrastructure deprivation, stress, lack of social support, and so on. An established tradition in sociology successfully studies the impact of neighbourhood environment on sociological outcomes such as educational attainment and labour market opportunity by taking advantage of multilevel analysis. Epidemiology picks up this tradition and uses multilevel analysis to examine group level effects on individual health. Thus, although for long time the social and health sciences have trodden quite different and independent paths, a new perspective integrating both approaches is now seeming to emerge.

The case of epidemiology is of particular interest since it has been argued that an integration of the social science approach and the health science approach will get at a better understanding of causal relations (see Susser and Susser (1996); Vineis et al. (2004); Weed (2000) and references therein). Some time ago the same methodological turn was also advanced in demography (Mosley and Chen, 1984) for studying the phenomenon of child

mortality in developing countries: social and economic determinants exert an impact on mortality through biological mechanisms. Hence, Mosley and Chen's approach—which now constitutes the received view in the specific field of child mortality in developing countries—incorporates both social and biological variables and integrates research methods proper to the two domains. Because of this move to consider both biological and social factors when explaining a given phenomenon, we see epidemiology as paradigmatic of the social and health sciences.

Epidemiology, being interested in the biological and social determinants of diseases and in their distribution on the population, has cognitive and practical goals. The cognitive goal concerns the aetiology of diseases, and the practical goal concerns the implications of such causal knowledge on policy making and also on causal attribution and diagnosis in particular individuals.

In §3 we shall focus our attention on the case of cancer epidemiology. This will help us make two major points in the paper. First, epidemiology is concerned with two different types of causal inferences. One is *generic* and concerns the population as a whole, and the other is *single-case* and concerns particular individuals.[2] Second, both causal inferences are essentially probabilistic; consequently we raise the question of which interpretation of probability best fits these probabilistic inferences. We shall argue that probabilities have to be interpreted according to an objective Bayesian approach in the single-case and according to a frequentist approach in generic inferences.

3 Case Study: Cancer

In the late '50s, when he was a scientific consultant for the Tobacco Manufacturers Standing Committee, Sir Ronald Fisher advanced that the correlation between smoking and lung cancer was due to an unknown genotype influencing both smoking behaviour and the predisposition to lung cancer, thus casting doubt on the hypothesis of a direct causal link from smoking to lung cancer.[3] The primary intent of Fisher was to point to the well known fact that correlation is not causation and that alternative explanations—

[2]There is a subtlety here: the distinction between population-level and individual-level (commonly drawn with regard to causal claims) does not quite correspond to the distinction (commonly drawn with regard to interpretations of probability) between the generic and the single case. This can be seen from the following claims. In general, inequality (population-level, generic) causes deterioration in health (population-level, generic). The inequality of her compatriots (population-level but single-case) is a cause of Naomi's deterioration in health (individual-level, single-case) (Glymour, 2003). To simplify matters we shall restrict our attention to the generic / single-case distinction in this paper.

[3](Fisher, 1957, 1958)

such as the one of an unknown genotype—could equally well account for statistical correlations.

Since then, cancer research has gone a long way. The link between smoking and cancer has been investigated under almost every possible angle. For instance, scientists try to establish the effect of tobacco control programs on declines in smoking and heart disease mortality (Barnoya and Glantz, 2004). In this type of study the causal effectiveness of tobacco consumption on lung cancer is implicitly assumed. However, if we can ascribe a decrease in mortality rates to a decrease in smoking, this will provide further epidemiological evidence. Fisher's hypothesis of a gene regulating cancer predisposition has not been dismissed, however. For instance, a recent study (Hwang et al., 2003) tried to assess the causal role of the p53 germline. Results indicate that a cancer predisposition due to a p53 mutation is significantly increased by cigarette smoking.

Specific studies concerning particular populations or subpopulations, such as those mentioned above, contribute to the mapping of the aetiology of cancer. From an epistemological viewpoint, the problem is how to gather together knowledge acquired in those specific studies in order to attain general epidemiological knowledge abut cancer. Consider for instance the two following papers: Lagiou et al. (2005); Vineis et al. (2004). These articles intend to provide a rather complete overview of the present state of affairs. That is, the aim is to summarize various epidemiological evidence on tobacco and cancer coming from specialised studies.

The mapping of the aetiology of cancer bears on several questions. One is the question of which types of cancer are due to tobacco consumption. There are several: lung, nasal cavity, stomach, liver, kidney, uterine cervix.[4] Another is the question of which biological factors are carcinogenic (e.g. hepatitis B and C virus, helicobacter pylori, human papilloma viruses, etc.) and which cancer sites are associated with these factors. A third question concerns the occupational chemicals that produce cancer; for instance, arsenic, asbestos, benzene, hair dyes, painting materials, soot. From a medical viewpoint it is important to understand which cancer sites are most often associated with exposure to these substances; on the other hand from a demographic viewpoint it is also important to figure out which parts of the population, in terms of social class or occupation, are more exposed to those carcinogenic substances and what the intensity of exposure is. A large part of medical and epidemiological research in cancer aetiology also focusses on

[4] Although epidemiological studies have variably shown positive, inverse or null associations between cigarette smoking and breast cancer, experimental studies indicate that tobacco smoke contains potential human breast carcinogens. See Terry and Rohan (2002).

the genetic factors predisposing or preventing cancer. Dietary behaviour, lifestyle factors (e.g., passive smoking, consumption of alcoholic beverages, ultraviolet radiation), and socio-demographic characteristics related to cancer are likewise investigated.

It is worth pointing out that results of particular studies do not automatically count as epidemiological evidence. Their soundness is evaluated according to specific criteria. The International Agency for Research on Cancer (IARC) classifies evidence of carcinogenicity into four categories:[5] sufficient, limited, inadequate and evidence suggesting lack of carcinogenicity. We have sufficient evidence when a positive relationship has been observed between the exposure and cancer in studies in which chance, bias and confounding could be ruled out with reasonable confidence. Evidence is limited if the working group considers the association credible, but chance, bias and confounding could not be ruled out with reasonable confidence. Evidence is inadequate if the available studies are of insufficient quality, consistency or statistical power to permit a conclusion about the presence or absence of a causal relation. Finally, evidence suggests lack of carcinogenicity when several studies are consistent in not showing a positive association between exposure to the agent and any studied cancer at any observed level of exposure.[6]

If this tells us something about the kind of evidence (biological and social) used to support causal statements and about the criteria used for evaluating those results, it doesn't say anything about the very concept of causation underlying causal analysis in cancer epidemiology.

In the following, we shall not review the extensive epidemiological literature to come up with an inventory of different concepts of cause. Such an overview of different concepts of cause in epidemiology is offered in recent publications (see Parascandola and Weed (2001) and references therein); virtues, faults and applicability of various concepts—production, necessary and sufficient, sufficient-component, counterfactual, probabilistic—are presented there. We won't even enter the debate about whether a probabilistic concept of causality does a better job than a deterministic one (although this seems to be the most recent point of view emerging, see Parascandola and Weed (2001); Vineis et al. (2004)). Instead, we shall focus on causal models and inferences in cancer epidemiology where, as matter of fact, causal relations are probabilistically characterized. It is worth noting, however, that a probabilistic methodology and epistemology do not necessarily im-

[5] See Vineis et al. (2004, p. 100) and Lagiou et al. (2005, p. 569), and the IARC web site http://monographs.iarc.fr/ENG/Preamble/index.php.

[6] Of course, evidence of lack of carcinogenicity requires that studies meet to a sufficient degree the standards of design, and in particular that bias, confounding and missclassification be ruled out with a reasonable degree of certainty.

ply that the *concept* of cause—i.e. the cause in the metaphysical sense—is itself probabilistic.[7] Consequently, "probabilistic" has to be understood in a non-metaphysical sense—i.e. as just referring to the use of probabilistic models for causal analysis in cancer epidemiology.

As mentioned above, there is now unquestionable evidence that tobacco is a powerful carcinogenic substance that can cause cancer in many different organs. It is also commonly agreed that while tobacco consumption raises the probability of developing cancer, tobacco consumption—whether active or passive—is not a sufficient cause. In other words, what scientists seek to establish is the extent to which smoking increases the probability of developing cancer, or the extent to which exposure to a carcinogenic substance such as asbestos influences cancer rates, or the extent to which particular dietary habits prevent—i.e. lower the probability of developing—cancer. Thus, causation of cancer is conceptualized in a probabilistic sense involving statistical terms and procedures. Whether studies are experimental or observational, the goal in both cases is to reduce uncertainty, by performing as many studies as possible to generate sensible summary statistics, and by reducing confounding and bias. Different statistical models, ranging from multiple regression analysis to structural modelling, are used to accomplish this task.

It is not hard to see that these probabilistic models only allow for probabilistic inferences. But what types of inferences are we concerned with? The first type of inference, which we shall call *generic* inference, aims at establishing whether or not a factor is a cause of disease by deciding whether, roughly, alterations in the frequency or intensity of this factor are accompanied by alterations in the frequency or intensity of disease. This corresponds to the naive causal statement 'smoking causes lung cancer.' The second type of inference, which we shall call *single-case* inference, is instead concerned with particular individuals. For instance, exposure to a known cause of cancer implies that this individual is now more likely to develop cancer. Assessing single-case probabilities to *particular* individuals is a real worry for practicians, and in fact one goal of Evidence Based Medicine is to provide guidelines to tackle particular situations.[8] Parascandola and Weed (2001, p. 908) echo Cox's, Holland's and Olsen's criticisms of probabilistic accounts: a probabilistic theory of causation, based on statistical inequalities, is inadequate since it leaves unclear what it means for smoking to raise the probability of an individual developing lung cancer. They argue that in this respect counterfactuals help. Instead, we argue that the problem is

[7] Parascandola and Weed (2001, p. 906) make this point but don't develop it further.

[8] See for instance online resources of the Centre for Health Evidence, http://www.cche.net/usersguides/.

not the inadequacy of probabilistic theories but rather a neglected aspect of them: the distinction between the generic and the single-case.

Lagiou et al. (2005, p. 569) seem to have grasped the importance of such a distinction, for they claim that although criteria such as those mentioned above are surely important for discerning causal association from non-causal association, they do not allow one to separate the different issues posed by (i) the results of a single study, (ii) the results of several studies, and (iii) the likelihood of causation in a particular individual. They acknowledge that an individual study (what they call level I) does not allow one to infer causation but can provide evidence when we already have several studies (level II) pointing to the same causal relation. Level II, however, is not sufficient to establish a causal link between an individual's exposure and disease. What it is still possible to do, they claim, is to infer from level II that the specific individual's illness was more likely than not caused by the specified exposure. This highlights the levels of causation and the different types of inferences we make either concerning the population as a whole or concerning particular individuals.

The discussion, however, is not pushed further. In particular, only to recognise different types of inferences is not enough: those inferences are probabilistic and consequently raise the problem of how probability has to be interpreted. This is the question we turn to next.[9]

4 Interpreting Probability

In the philosophy of probability there is a wide-ranging debate about interpretation. Indeed, discussions about the *meaning* of probability began as early as the first formulations of probability theory. We direct the reader to Hacking (1975); Gillies (2000) for interesting historical overviews.

As we have seen in §3, causal inferences in cancer epidemiology are essentially probabilistic. A standard objection to probabilistic theories of causation is that the claim 'smoking raises the probability of lung cancer' is ambiguous for it might be interpreted in different ways. It might say that within the population, the proportion of those who develop lung cancer is greater amongst smokers. It might also say that if a *particular patient* smokes, then it's more likely that she will develop lung cancer. This ambiguity motivates a distinction between different levels of causation and consequently between different types of causal inferences: generic and single-case. Those claims state, in different terms, a probabilistic relation between

[9]Inferences in cancer epidemiology, whether generic or single-case, are also based on the computation of relative risks and odds ratios. In the appendix we shall briefly address the problem of correctly interpreting these measures and argue that the generic/single-case distinction is again illuminating.

smoking and lung cancer. But what does probability mean in these cases? Is there a unique interpretation fitting the two examples equally well? Or should we attach a different meaning to probability depending on the claim at stake?

It is not hard to see that those claims state quite different things. A generic causal claim posits an average causal relation, which is supposedly valid for the majority of individuals exposed within the population. The adoption of a probabilistic framework entails, by itself, that such a causal claim is not a universal and necessary law, and, consequently, that not every individual in the population who instantiates the cause will instantiate the effect. In fact, as is well known, some smokers never develop cancer, and some non-smokers instead do. This leads us to ask what the meaning of a single-case causal claim is. Two meanings ought to be distinguished. In one case we wish to make a prediction: your smoking now makes you more likely to develop cancer in the future. A second interpretation is instead retrospective: it is likely that smoking caused you to develop lung cancer.

It is then apparent that the meaning of probability for generic causal claims is connected with frequency of occurrence, whereas for single-case causal claims the meaning is closer to something like belief or credence about what will happen or had happened. A competing interpretation for the single-case is single-case chance. However, as we shall see, single-case chance raises problems of epistemic access.

The overview we offer next is meant to see whether any of the leading interpretations provide sensible meaning both for generic and single-case claims. However, we shall see that none of these interpretation succeeds. In §5 we argue that a frequency interpretation is needed for the generic and that a Bayesian approach is instead needed for the single case. In §6 we will argue that practical considerations motivate choosing the objective Bayesian approach over the empirically-based approach.

We will now sketch very briefly the main features of the four leading contenders: (i) the classical and logical interpretations; (ii) the physical interpretations: frequency and propensity; (iii) the subjective interpretation; (iv) the objective Bayesian interpretation.

The classical interpretation of probability defines probability as the ratio between the number of favourable cases and the number of all equipossible cases. The easiest way to grasp the meaning of favourable and equipossible cases is to think of dice. The six sides of a die constitute the probability space—i.e. the six possible outcomes. Assuming that the die is not biased, the six sides are all equipossible—i.e. they all have the same probability of occurring uppermost. The favourable cases constitute the event we are interested in. For instance, the probability that an even number will result

is given by the number of favourable cases (sides 2, 4 and 6) over the total number of the equipossible cases (for an unbiased die: 1, 2, 3, 4, 5 and 6) which gives $\frac{3}{6}$, i.e. $\frac{1}{2}$.

First developed by Laplace (1814), a similar interpretation was also proposed by Pascal. Probability values are assigned in the absence of any evidence—the probability of an event is simply the fraction of the total number of possibilities in which the event occurs. The notion of equipossibility is expressed by what Keynes called the Principle of Indifference. This principle states that whenever there is no evidence favouring one possibility over another, these possibilities have the same probability. The classical interpretation seems especially well suited to games of chance, although it is sometimes objected that this interpretation suffers the problem of circularity, for equipossible means equiprobable, hence 'probable' is not properly explicated. A second traditional objection is that the classical interpretation is of scarce applicability in science; in fact, adopting the classical interpretation we have no meaningful way to express knowledge of the population probabilistically, nor to evaluate individual hypotheses.

A generalization of the classical interpretation is the so-called logical interpretation, advanced by Keynes (1921); Jeffreys (1939); Carnap (1950). This interpretation depends on the Principle of Indifference and thus rests on the idea that probabilities can be determined a priori by an examination of the space of possibilities, but only when no knowledge indicating unequal probabilities is available. The main aim of the logical interpretation is to provide an account, as general as possible, of the degree of support or confirmation that a piece of evidence e confers upon a given hypothesis h. In Carnap's notation, the c-function precisely expresses this idea.[10]

According to the physical view, probability values are quantitative expressions of some feature of the world, not of our knowledge or beliefs. The physical view is typically taken to encompass the frequency and the propensity interpretations . A simple version of frequentism , due to Venn (1866), states that the probability of an attribute A in a finite reference class B is the relative frequency of the actual occurrence of A within B. Further developments of frequentism are due to von Mises (1928) and Reichenbach (1935), who consider infinite reference classes and identify probabilities with the limiting relative frequencies of events or attributes therein. This second sort of frequentism is also advocated by Salmon (1967). Limiting relative

[10]As well as appealing to symmetry or indifference, Carnap (1950, §§41–42) bases probability values upon knowledge of physical probabilities. He says, 'in these cases the probability is determined with the help of a given frequency and its value is either equal or close to that of the frequency' (Carnap, 1950, §42B). Consequently, Carnap's development of the logical interpretation might be classified alongside the objective Bayesian approach, which we will shortly introduce.

frequencies serve, in his approach, to determine the probability of evidence. Note that frequentism interprets generic probabilities; in order to associate a frequency with a single case, a unique reference class must be associated with the single case; that this can not be done in general is known as the *reference class problem*.

The propensity interpretation is also located in the physical realm, since probability is 'in the world', so to speak, rather than 'in our heads'—as it is in the subjectivist approach and the classical and logical interpretations. Probability is here conceived as a physical propensity, or disposition, or tendency, of a given type of physical situation to yield an outcome of a certain kind, or to yield a long run relative frequency of such an outcome. The propensity interpretation was advanced by Popper (1957, 1959), who was motivated by the desire to make sense of single-case probabilities, for instance in quantum mechanics.

In the subjective interpretation probabilities are quantitative expressions of an agent's opinion, or degree of belief, or epistemic attitude, or something similar. First advances are due to Ramsey (1926); de Finetti (1937). De Finetti's viewpoint is paradigmatic of personalistic approaches, for he firmly stated that probability does not exist (in the physical sense), and that it is possible to reconstruct and deduce probability theory just relying on the subjectivist interpretation (de Finetti, 1993, pp. 248 ff). In subjectivist approaches, also called subjective Bayesian, probabilities are analyzed in terms of betting behaviour. Probabilities are identified with the announcement of the betting odds that a rational agent is willing to accept. A Dutch book (against an agent) is a series of bets, each acceptable to the agent, but which collectively guarantee her loss, whatever happens. Two Dutch book theorems then follow. (i) If an agent's degrees of belief violate the probability calculus, then she is liable to a Dutch book, and, conversely, (ii) if an agent's degrees of belief conform to the probability calculus, then no Dutch book can be made against her. A series of bets is called *coherent* if it is not susceptible to a Dutch book. In subjectivist approaches, obedience to the probability calculus is taken to be a necessary and sufficient condition for rationality.

It is typically objected that this personalistic account leads to arbitrariness, that is, it is too subjective. In fact, two agents with exactly the same evidence may assign different probability values to the same event and be equally rational, provided that they do not violate the probability calculus. (It is worth pointing out that de Finetti's betting interpretation derives probabilities from utilities and rational preferences. The intimate link between utility of outcomes and probabilities is even more prominent in the approaches of Savage (1954); Jeffrey (1965). The main idea of the utility

interpretation is that probabilities and utilities can be derived from preferences among options that are constrained by certain putative consistency principles.)

A solution to the objection of arbitrariness is attempted by empirically-based subjectivism and also by the objective Bayesian interpretation of probability. These interpretations require that further constraints beyond coherence be satisfied before an agent's degrees of belief can be considered rational. The former approach, empirically-based subjectivism, was advocated e.g. by Salmon (1967, 1990). This approach requires that empirical information constrain degrees of belief: if, for instance, an agent knows that 60% of people with a certain type of cancer recover, knows that a particular patient has this type of cancer, but knows nothing else pertinent, then she should believe that the patient will recover to degree 0.6. The latter approach, objective Bayesianism, was put forward by Jaynes (1957) and goes beyond empirically-based subjectivism. According to this view, lack of knowledge should also constrain degrees of belief: in the absence of evidence the agent should be as equivocal as possible, e.g., if the agent does not have any knowledge at all pertinent to a cancer patient then she should believe that the patient will recover to degree 0.5; if her knowledge constrains her degree of belief to fall in the interval $[0.6, 0.8]$ then she should chose the point that most equivocates between recovery and non-recovery, i.e., 0.6.[11] Thus both information and lack of information about the world should be taken into account in shaping epistemic probabilities. Information-theoretic considerations motivate the use of entropy as a measure of the extent to which a probability function equivocates; consequently Jaynes put forward the *maximum entropy principle*, which provides a formal framework for objective Bayesianism: the agent's belief function should be a probability function, from all those that satisfy constraints imposed by evidence, that has maximum entropy. In this framework, on a finite domain an agent's background knowledge fully determines the degrees of belief that she ought to adopt.[12]

Note that degree-of-belief interpretations—including subjectivism, empirically-based subjectivism, and objective Bayesianism—interpret single-case rather than generic probabilities: degrees of belief are associated with bet-

[11] Note that the interval $[0.6, 0.8]$ is not a confidence interval. If a study indicates that 70% of people recover and provides a confidence interval $[0.6, 0.8]$, then the best (albeit defeasible) evidence is that 70% of people recover and the agent should simply set her degree of belief in recovery to 0.7. Rather, the interval constraint might be generated by two studies, one of which finds a 60% recovery rate, the other of which finds 80%, and neither of which is to be preferred over the other (on the grounds of sample size, specificity etc.)—then it is reasonable to place one's degree of belief somewhere in the ordered interval generated by the frequencies.

[12] (Williamson, 2005, §5.3). Note that this is not necessarily the case on infinite domains—see Williamson (2006c, §19).

ting, and a bet in a generic outcome makes little sense.

In §5 we will argue that generic causal claims demand a frequency interpretation, while single-case claims require an empirically-based or objective Bayesian interpretation. In §6 will go further in arguing that an objective Bayesian interpretation should be chosen.

5 Desiderata

In this section we shall put forward some requirements that an interpretation of probability should meet.

A philosophical theory of probability should:

Objectivity: account for the objectivity of probability,

Calculi: explain how we reason about probability,

Epistemology: explain how we can know about probability,

Variety: cope with the full variety of probabilistic claims that we make,

Parsimony: be ontologically parsimonious.

We shall discuss each of these desiderata in turn, paying special attention to the application to causal models of cancer.[13]

Objectivity

Many applications of probability invoke a notion of probability that is objective in a logical sense: there is a fact of the matter as to what the probabilities are; if two agents disagree about a probability, at least one of them must be wrong.[14] For example, the probability that a patient's breast cancer will recur after treatment is supposed to depend on features of the cancer (e.g. whether it is metastatic, whether it is HER2 positive, its ER status), of the treatment, and of the patient. It is not simply a matter of personal opinion: if two prognostic probabilities differ, at least one of them must be wrong. A philosophical theory of probability should yield a notion of probability that is objective in this logical sense—otherwise it is not meeting the demands of these particular applications.

Clearly the subjective interpretation of probability suffers in this respect. According to the subjective theory, probabilities are degrees of belief and one can adopt any prior probabilities one likes as one's degrees of belief. According to the subjective theory, then, one agent can give probability

[13] See Williamson (2006a) for discussion of similar desiderata as requirements of a philosophical theory of causality.

[14] Logical objectivity contrasts with the ontological sense of objectivity: probabilities are ontologically objective if they exist as physical entities.

0.9 to the patient's breast cancer recurring, another agent with the same knowledge of the situation can give probability 0 to the same event, and neither agent can be considered wrong.[15] Empirically-based subjectivism also suffers, but to a lesser extent: if frequencies are known then probability assignments are not arbitrary, but where frequencies are not known an agent can choose her degrees of belief arbitrarily. The classical and logical interpretations can also suffer at the hands of Objectivity, since different agents can construe different partitions of events as equipossible.

In contrast, frequency, propensity and objective Bayesian interpretations all yield objective probabilities of varying forms. A frequency is objectively determined by a reference class; a propensity is objectively determined by the history of the universe up to the present time; under objective Bayesianism a probability is objectively determined by an agent's knowledge. Thus these interpretations fare better with respect to this desideratum.

Calculi

Probabilities are manipulated and inferences are drawn from them by means of the probability calculus. This mathematical apparatus, based on axioms put forward by Kolmogorov (1933), has by now become well entrenched. Consequently a philosophical theory of probability should yield a notion that satisfies the axioms of probability. Otherwise it is not a theory of probability, but a theory of something else.

Some theories suffer in this respect. According to some accounts, probabilities are not real numbers but are intervals of numbers, pairs of real numbers, or qualitative entities.[16] According to other accounts probabilities satisfy some axioms but not others—the frequency theory of von Mises (1928), for instance, does not satisfy the axiom of countable additivity; the propensity theory has problems with conditional probabilities.[17] With respect to this desideratum, then, degree-of-belief interpretations (subjectivism, empirically-based subjectivism, and objective Bayesianism) fare better than these other approaches.

[15] Proponents of the subjective account tend to respond in two ways: by saying that the subjective theory can account for objectivity in the long run as different agents' beliefs converge to frequencies, and by saying that there is no further objectivity to be found. While the first claim is notoriously problematic (Williamson, 2005, §2.8), the second claim is simply dangerous. If the subjectivist has no knowledge that bears on recurrence of breast cancer and awards a degree of belief 0.9, instead of the objectively-determined middling value 0.5, then she may initiate unnecessarily aggressive treatment rather than collect further evidence—see §6. Thus there is further objectivity—derived from the need to equivocate in the absence of evidence—that the subjective account ignores.

[16] See e.g. Keynes (1921); Kyburg Jr (2003); Walley (1991).

[17] (Humphreys, 2004)

Epistemology

We come to know about probabilities in various ways: we measure population frequencies, we appeal to symmetry arguments or scientific theories, we make educated guesses, we derive some probabilities from others using the probability calculus. A philosophical theory of probability should explain how we can use such techniques to discover probabilities. If the theory rejects some of these techniques it should say where they go wrong and why they are apparently successful.

This desideratum is a stumbling block for several theories. The classical, logical and subjective theories can not account for the widespread use of frequencies, while the frequency theory can not explain how degrees of belief can offer access to probabilities. The propensity theory is oft criticised for being metaphysical: it connects probability with scientific theories and even degrees of belief,[18] but struggles to identify a precise link with frequency. However the empirically-based subjectivist and objective Bayesian approaches allow background knowledge of any form—frequencies, symmetries, scientific theories included—to constrain an agent's rational degrees of belief; by design these theories admit a variety of sources of probabilistic knowledge.

Variety

Probabilistic claims are extremely varied. For instance, claims are made about single-case probabilities (e.g. the probability that a particular patient's cancer will recur) and generic probabilities (e.g. the probability of recurrence among those who receive radiotherapy). Moreover, probabilities are attached to a variety of entities, including events, sets, variables, sentences, propositions and hypotheses. A philosophical theory of causality should be able to cope with this variety—it should account for each use of probability, or, if some uses are to be viewed as illegitimate, it should say how such uses should be eliminated in favour of the legitimate uses. Otherwise, the theory is at best a partial theory, a theory of *some* of the uses of probability.

This desideratum is a problem for many of the interpretations of probability. The frequency and propensity theories can not ascribe a probability to a given hypothesis, but only yield the probability of observing a sample if the hypothesis is true—on this point see Courgeau (2004). Moreover, the frequency theory is a generic theory; it views single-case probabilities as illegitimate but provides no means of interpreting single-case claims in terms of frequencies. If single-case probabilities are to be abandoned the theory really ought to explain why their use, if so erroneous, is apparently

[18] (Lewis, 1980)

so successful. Other interpretations ascribe probabilities to single cases, and do not provide a means for interpreting population-level probabilities. It seems that pluralism is the only option: have one interpretation for the single-case and another for generic claims. But then work needs to be done to explain why we apparently have a single concept of probability when in fact there are at least two. The empirically-based subjectivist or the objective Bayesian route is perhaps most attractive here: use frequencies in the generic case, and use these frequencies to constrain single-case degrees of belief. The two notions of probability, frequency and degree of belief, are tightly connected under these accounts and do not seem so disparate after all.

Parsimony

Arguably a philosophical theory of probability should not make unwarranted ontological commitments: if one can reduce probabilities to something else in one's ontology then one should do that rather than take probabilities as primitive. This is just Ockham's razor; it may be viewed as a methodological or psychological requirement and as such subsidiary to the other desiderata.

Parsimony tells against the propensity interpretation, which usually takes probabilities to be primitive. Keynes (1921) in his development of the logical view also takes probabilities to be primitive. In contrast a frequency is a feature of a sequence of observed outcomes and so presumably reducible to entities already in a natural ontology. Similarly rational degrees of belief, the entities of the subjective, the empirically-based subjective and the objective Bayesian interpretations, will already be included in an ontology and do not count as an ontological extra. Of course all this depends on ontology; an ontology that contains only propensities may be more parsimonious than one that contains rational degrees of belief among other things.

We see then that these desiderata help to isolate a viable interpretation of probability. The propensity theory falls foul of Calculi, Epistemology, Variety and Parsimony; the frequency theory of Calculi, Epistemology and Variety; the classical, logical and subjective theories of Objectivity, Epistemology and Variety. The empirically-based subjective theory does well, though perhaps suffers with respect to Objectivity. The objective Bayesian interpretation seems to offer the most promise, when twinned with a frequency interpretation of generic probabilities. This combination is a particularly attractive way of interpreting probability in causal models for cancer: crucially, perhaps, the epistemology desideratum is satisfied—we can know about generic probabilities as well as single-case probabilities; moreover this

combination makes sense both of generic causal inferences where probabilities can be interpreted as frequencies, and of single-case causal inferences where objective Bayesian probabilities are determined by empirical knowledge including frequencies.

6 Objective Bayesianism

Clearly frequencies are required to make sense of generic probabilities, e.g. the probability of surviving more than five years given metastatic breast cancer is 0.4. We have also suggested that degrees of belief constrained by frequencies should be used to make sense of the single case: if one knows only the aforementioned generic probability then one should believe that Audrey, who has metastatic breast cancer, will survive more than five years, to extent 0.4. This ties the two levels together in a natural way: generic knowledge yields predictions about the single case.

As yet though, this leaves two interpretations of probability for the single case. First, we have empirically-based subjectivism: an agent's degrees of belief ought to be constrained by knowledge of frequencies; in the absence of such knowledge they may be chosen arbitrarily. The second alternative is objective Bayesianism: an agent's degrees of belief ought to be constrained by knowledge of frequencies; in the absence of such knowledge they should be as equivocal as possible. In our view the latter approach should be adopted, as we shall now explain.

The main reason for preferring objective Bayesianism over the empirically-based subjective theory is that objective Bayesianism is on average more cautious when it comes to risky decisions. In cancer applications, single-case probabilities are used to make treatment decisions.[19] For example, if the probability of recurrence in a particular patient is very high, aggressive treatments might be used; if the probability of recurrence is very low then no further treatment is given; otherwise more evidence is garnered and non-aggressive treatments are given. Now suppose empirical evidence forces degree of belief in recurrence to lie between 0 and 0.4, say. Under the empirically-based subjective theory, an agent is free to choose any degree of belief within this interval $[0, 0.4]$. So the agent may set degree of belief 0, which will trigger abandonment of treatment. But under objective Bayesianism, the agent must choose the most equivocal—i.e. middling—degree of belief from this interval. So she must have degree of belief 0.4, which may trigger the collection of more evidence in order to reach a firmer opinion, and may trigger a non-aggressive treatment in the meantime. In general, high-risk decisions tend to be triggered by high or low degrees of belief; the objective Bayesian protocol ensures that such drastic actions only

[19] See e.g. Williams and Williamson (2006).

get taken if there is sufficient empirical evidence to force the extreme probabilities required to trigger them. If evidence is lacking then more middling probabilities must be adopted so that less risky actions can be initiated and further evidence can be collected.[20]

Tim McGrew has posed the following objection to this argument:

> Suppose that someone was diagnosed with cancer, underwent treatment T_1, and is being assessed for the success of T_1. Unless there is good evidence that T_1 was successful, a more radical treatment T_2 is indicated. Empirical information indicates a probability interval $[0.6, 0.95]$ for the success of T_1. It would appear that objective Bayesianism requires us to adopt the probability 0.6 that T_1 was successful, which may lie within the zone that triggers aggressive treatment T_2. There may be a good answer to this worry, but it does, *prima facie*, cast doubt on the idea that objective Bayesianism always errs on the side of caution, where caution can be equated with a preference for the less radical treatment over the more radical.[21]

In response, the first thing to note is that the most equivocal probability, in this case 0.6, will not always be the most cautious—this depends on the particular decision structure. At best we can say that *on average* with respect to risky decisions the maximum entropy approach is the more cautious policy—see Williamson (2006b, §8) for a full discussion of this point. We can see that equivocal probabilities tend to be more cautious when we see how decision scenarios arise. Consider the above scenario. Here the decision rule is that T_2 will be instigated unless the probability of the success of T_1 is greater than some threshold. Now a middling probability such as 0.6 indicates a lack of evidence and will normally trigger the collection of further evidence to try to shift the probability to one extreme or the other. In this case, however, collecting further evidence is not an option; this suggests that the costs and risks associated with collecting further evidence are greater than those associated with the treatment T_2, for otherwise the decision structure would be different. T_2 is triggered by 0.6 because, aggressive though it is, this treatment is the least risky action available. So even in this example, the objective Bayesian degree of belief is most cautious. The fact is that when we set up decision protocols, the risky actions tend to be triggered by extreme probabilities and conversely equivocal probabilities tend to trigger the less risky actions; if a middling degree of belief triggers a

[20] This argument is presented in detail in Williamson (2006b).

[21] Personal communication. We are very grateful to Tim McGrew for this point and other insightful comments.

risky action then that is because it is the least risky action overall. We must not equate caution with a preference for the less radical treatment over the more radical, but with the choice of a less risky action over a more risky action (in the absence of evidence that warrants the more risky action). Objective Bayesianism is then the more cautious policy, on average.

In sum, when we take the applications of single-case probabilities into account—in particular the application to cancer treatment—it becomes clear that these probabilities are a guide to action. Some actions are more drastic than others, and objective Bayesianism ensures that such actions are not embarked upon lightly.

7 Conclusion

In this paper we have argued that epidemiology can be seen as paradigmatic of the methodology of the social and health sciences. As far as causal analysis is concerned, we considered cancer epidemiology for two reasons: causal models for cancer (i) try to take into account both socio-economic and biological factors, and (ii) are essentially probabilistic.

The analysis of causal models and of the probabilistic inferences they induce suggests that two levels of inference ought to be distinguished: generic and single-case. Generic causal statements aim at positing average causal relations, for instance by claiming that tobacco consumption is a powerful carcinogen or that healthy dietary habits can possibly prevent cancer. Single-case causal statements tend to concern, instead, particular individuals. They are used for diagnosis or for causal attribution, for instance to assess the probability of recurrence of breast cancer in a particular patient.

As a matter of fact, both generic and single-case inferences are probabilistic and therefore raise the problem of how probability should be interpreted. An overview of the leading interpretations—classical and logical, frequency, propensity, subjective, empirically-based subjective, and objective Bayesian—shows that if we want to make sense of probabilistic statements at both levels we have to opt for a pluralist interpretation: the frequency interpretation is most appropriate for generic causal claims, and the single case demands a Bayesian interpretation in which probabilities are thought of as degrees of belief constrained by frequencies.

Moreover, cancer epidemiology appears to be of particular interest because it is both concerned with gaining general causal knowledge—e.g. about cancer aetiology—and with applying such general knowledge to a particular individual. This is not true of all social and health sciences. Many disciplines are more concerned with the general level and only indirectly with the single-case or the other way around. Demography, for instance, studies migration behaviour of populations but it is not directly interested in the

probability that a particular individual will migrate. On the other hand, it is oft said that in medicine there are not illnesses but only ill persons to cure. Physicians, then, are more concerned with single-case probabilities rather then with frequencies of disease.

Though not directly concerned with individuals, social sciences such as sociology, demography or economics do have a bearing on the individual since their results orient and guide public policies, for instance to reduce unemployment or to discourage tobacco consumption. On the other hand, to correctly assign single-case probabilities, physicians do need to take generic probabilities into account. Thus, we'd better have a unified account of the interpretation of probability that makes sense both at the generic level and at the single-case. Such an account—we have argued—is objective Bayesianism twinned with the frequency interpretation.

Acknowledgements

This research forms a part of the project *Causality and probability in the social and health sciences*. We are grateful to the British Academy and to the FSR (Université catholique de Louvain) for funding this project. We are also grateful to two referees for very helpful comments.

Federica Russo
Philosophy, SECL, University of Kent at Canterbury, UK.
f.russo@kent.ac.uk

Jon Williamson
Philosophy, SECL, University of Kent at Canterbury, UK.
j.williamson@kent.ac.uk

A Probabilities, Odds and Risks

The medical and epidemiological sciences often summarise results—e.g. of logistic regressions or of meta-analyses—by means of risks and odds. Although these are associational measures, arguably they have causal import insofar as they provide evidence for a generic causal claim—e.g. smoking causes lung cancer—or inform single-case inferences—e.g. predicting the survival time of a particular individual—on the basis of these measures.

In several types of biomedical research, for instance case-control, cohort, cross-sectional or experimental studies, risks and odds are used to quantify the strength of the relation between two binary variables: a particular outcome (disease) and presence of factor (exposure). Those results are customarily presented in 2x2 contingency tables.

Let E and D denote two binary or dichotomous variables, each having only two possible levels. For the explanatory variable (exposure) E: *exposed, unexposed*, and for the outcome (disease) D: *yes, no*. Results in the table are presented as counts of observations at each level. A 2x2 contingency table has thus 4 cells. The table below shows the general layout of a 2x2 contingency table.

	Disease	
Exposure	Yes	No
Exposed	n_{11}	n_{12}
	p_{11}	p_{12}
Unexposed	n_{21}	n_{22}
	p_{21}	p_{22}

The notation n_{ij} refers to the *number* of subjects observed in the corresponding cell, i.e. to the number of observations in the i-th row ($i = 1, 2$) and j-th column ($j = 1, 2$); the notation p_{ij} refers to the *proportion* of subjects observed in the corresponding cell, where $p_{ij} = n_{ij}/n$, n being the total number of observed subjects. The notation $P(E)$ and $P(D)$ will refer to the marginal probabilities of exposure and disease, respectively.

With this data we can compute relative risks, odds, odds ratios and estimate probabilities.

Relative risk

The relative risk (RR) is defined as the ratio of risk in the exposed and unexposed group:

$$\frac{n_{11}/n}{n_{21}/n} = \frac{p_{11}}{p_{21}}$$

Thus RR measures the *incidence* of the disease. $RR > 1.0$ indicates that the risk of disease is increased when the risk factor (exposure) is present; $RR < 1.0$ indicates that the risk of disease is decreased when the risk factor is present, i.e. the factor is a protective factor or preventative. RR can also be given a definition in terms of conditional probabilities:

$$\frac{P(D|E)}{P(D|\neg E)}.$$

Odds ratio

The odds ratio (OR) is another way to compare proportions in a 2x2 contingency table. OR is computed from odds: it is the ratio of the odds of

disease in the exposed group and the odds of disease in the unexposed group. The odds of an outcome is equal to the probability that the outcome does occur divided by the probability that the outcome does not occur. In a 2x2 contingency table, the probability of an outcome is equal to the number of times the outcome is observed divided by the total observations.

$$OR = \frac{Odds_{exp}}{Odds_{unexp}}$$

$$Odds_{exp} = \frac{n_{11}/(n_{11}+n_{12})}{n_{12}/(n_{11}+n_{12})} = \frac{n_{11}}{n_{12}}$$

where $n_{11}/(n_{11}+n_{12})$ is the probability that the disease occurs in the exposed group and $n_{12}/(n_{11}+n_{12})$ is the probability that the disease does not occur in the exposed group. In terms of conditional probabilities,

$$Odds_{exp} = \frac{P(D|E)}{P(\neg D|E)}$$

.
Similarly,

$$Odds_{unexp} = \frac{n_{21}/(n_{21}+n_{22})}{n_{22}/(n_{21}+n_{22})} = \frac{n_{21}}{n_{22}},$$

where $n_{21}/(n_{21}+n_{22})$ is the probability that the disease occurs in the unexposed group and $n_{22}/(n_{21}+n_{22})$ is the probability that disease does not occur in the unexposed group. Again, in terms of conditional probabilities,

$$Odds_{unexp} = \frac{P(D|\neg E)}{P(\neg D|\neg E)}$$

OR can now be computed as

$$\frac{n_{11}/n_{21}}{n_{12}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}.$$

or, equivalently

$$\frac{P(D|E)}{P(\neg D|E)} \times \frac{P(\neg D|\neg E)}{P(D|\neg E)}.$$

It is worth noting that the odds ratio of exposure $OR = \frac{Odds_{exp}}{Odds_{unexp}}$ is equal to the odds ratio of disease $OR = \frac{Odds_{Dyes}}{Odds_{Dno}}$.There is a mathematical relation between odds and probabilities:

$$P(D|E) = \frac{Odds_{exp}}{1 + Odds_{exp}}$$

and

$$Odds_{exp} = \frac{P(D|E)}{1 - P(D|E)}.$$

Interpreting *RR*, *OR* and probabilities

Although calculations are fairly easy, the interpretation appears to be more tricky. For instance, Sistrom and Garvan (2004, p. 16) claim: '... an RR equal to 2.0 means that an exposed person is twice as likely to have an adverse outcome as one who is not exposed, and an RR of 0.5 means that an exposed person is half as likely to have the outcome.' Or (ibidem): 'Odds and probabilities are different ways of expressing the chance that an outcome may occur.' Here there is a tension between a generic and single-case interpretation of RR, OR and probabilities.

Similarly, Bland and Altman (2000) on the one hand explain the odds and odds ratios as means to compare *groups*, but, on the other, in giving an example, they talk in terms of individuals: 'The probability that a child with eczema will also have hay fever is estimated by the proportion 141/561 (25.1%). ... Similarly, for children without eczema the probability of having hay fever is estimated by 928/14453 (6.4%).'

Two remarks are in order. Firstly, this last quotation may look puzzling unless we make clear that the 'child' that Bland and Altman refer to is a *statistical individual*, i.e. an individual randomly sampled from the population. In this case the probability may be construed as generic rather than single-case.

Secondly, as shown above, the calculation of risks and odds involves *proportions*—i.e. the numbers of subjects who got/didn't get the disease and were/were not exposed to the factor compared to the whole population. Because calculation involves proportions, RR and OR have a natural *generic* interpretation and do not make sense in the single case. Consequently, the corresponding probabilities need a frequentist interpretation. However, if the definitions in terms of conditional probabilities are preferred instead, one might argue that these probabilities are not the frequencies drawn from the 2x2 table but subjective probabilities. If so, then risks and odds are all single-case, referring to a single individual, and they do not say anything about the population. But of course this view is prone to the objections raised in §5. From a normative point of view, rational degrees of belief should be based on empirical evidence such as frequencies, and otherwise maximally equivocal. Thus an *objective Bayesian* interpretation ought to be preferred instead for the single case. A further advantage of the objective Bayesian interpretation is that if a contingency table based on frequencies is incomplete, one can generate a complete contingency table by filling in missing values with objective Bayesian probabilities.

BIBLIOGRAPHY

Barnoya, J. and Glantz, S. (2004). Association of the tobacco control program with declines in lung cancer incidence. *Cancer Causes and Control*, 15:689–695.

Bland, J. M. and Altman, D. G. (2000). The odds ratio. *British Medical Journal*, 320:1468.

Carnap, R. (1950). *Logical foundations of probability*. Routledge and Kegan Paul, London.

Courgeau, D. (2004). Probabilité, démographie et sciences sociales. *Mathematics and Social Sciences*, 167:27–50.

de Finetti, B. (1937). Foresight. its logical laws, its subjective sources. In Kyburg, H. E. and Smokler, H. E., editors, *Studies in subjective probability*, pages 53–118. Robert E. Krieger Publishing Company, Huntington, New York, second (1980) edition.

de Finetti, B. (1993). *Induction and Probability*. D. Montanari and D. Cocchi, editors. CLUEB, Bologna.

Fisher, R. (1957). Alleged dangers of cigarette smoking. *British Medical Journal*, 2:297–298.

Fisher, R. (1958). Lung cancer and cigarettes. *Nature*, 182.

Gillies, D. (2000). *Philosophical theories of probability*. Routledge, London and New York.

Glymour, B. (2003). On the metaphysics of probabilistic causation: lessons from social epidemiology. *Philosophy of Science*, 70:1413–1423.

Hacking, I. (1975). *The emergence of probability*. Cambridge University Press, Cambridge.

Humphreys, P. (2004). Some considerations on conditional chances. *British Journal for the Philosophy of Science*, 55:667–680.

Hwang, S.-J., Shu-Chung, C. L., Lonzano, G., Amos, C., Gu, X., and Strong, L. (2003). Lung cancer risk in germline p53 mutation carriers: association between an inherited cancer predisposition, cigarette smoking, and cancer risk. *Human Genetics*, 113:238–243.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.

Jeffrey, R. (1965). *The logic of decision*. University of Chicago Press, Chicago IL, second (1983) edition.

Jeffreys, H. (1939). *Theory of Probability*. Clarendon Press, Oxford, third (1961) edition.

Keynes, J. M. (1921). *A treatise on probability*. Macmillan (1948), London.

Koch, R. (1882). Über die ätiologie der Tuberkulose. In *Verhandlungen des Kongresses für Innere Medizin*, Erster Kongress, Wiesbaden.

Kolmogorov, A. N. (1933). *The foundations of the theory of probability*. Chelsea Publishing Company (1950), New York.

Kyburg Jr, H. E. (2003). Are there degrees of belief? *Journal of Applied Logic*, 1:139–149.

Lagiou, P., Adam, H.-O., and Trichopoulos, D. (2005). Causality in cancer epidemiology. *European Journal of Epidemiology*, 20:565–574.

Lewis, D. K. (1980). A subjectivist's guide to objective chance. In *Philosophical papers*, volume 2, pages 83–132. Oxford University Press (1986), Oxford.

Mosley, W. and Chen, L. (1984). An analytical framework for the study of child survival in developing countries. *Population and Development Review*, 10:25–45. Supplement: Child Survival: Strategies for Research.

Parascandola, M. and Weed, D. (2001). Causation in epidemiology. *Journal of Epidemiology and Community Health*, 55:905–912.

Pickett, K. and Pearl, M. (2001). Multilevel analysis of neighbourhood socioeconomic context and and health outcomes: a critical review. *Journal of Epidemiology and Community Health*, 55:111–122.

Popper, K. R. (1957). The propensity interpretation of the calculus of probability and the quantum theory. In Körner, S., editor, *Observation and Interpretation*, volume 9, pages 65–70. Butterworths.
Popper, K. R. (1959). The propensity interpretation of probability. *British Journal for the Philosophy of Science*, 10:25–42.
Ramsey, F. P. (1926). Truth and probability. In Kyburg, H. E. and Smokler, H. E., editors, *Studies in subjective probability*, pages 23–52. Robert E. Krieger Publishing Company, Huntington, New York, second (1980) edition.
Reichenbach, H. (1935). *The theory of probability: an inquiry into the logical and mathematical foundations of the calculus of probability*. University of California Press (1949), Berkeley and Los Angeles. Trans. Ernest H. Hutten and Maria Reichenbach.
Salmon, W. C. (1967). *Foundations of Scientific Inference*. University of Pittsburgh Press, Pittsburgh.
Salmon, W. C. (1990). Rationality and objectivity in science, or Tom Kuhn meets Tom Bayes. In Wade Savage, C., editor, *Scientific theories*, pages 175–204. University of Minnesota Press, Minneapolis. Minnesota Studies in the Philosophy of Science 14.
Savage, L. (1954). *The Foundations of Statistics*. John Wiley.
Sistrom, C. L. and Garvan, C. W. (2004). Proportions, odds, and risk. *Radiology*, 230:12–19.
Susser, M. and Susser, E. (1996). Choosing a future for epidemiology II: from black box to chinese box and ecoepidemiology. *American Journal of Public Health*, 86:674–677.
Terry, P. and Rohan, T. (2002). Cigarette smoking and the risk of breast cancer in women: a review of the literature. *Cancer Epidemiology, Biomarkers and Prevention*, 11:953–971.
Venn, J. (1866). *Logic of chance: an essay on the foundations and province of the theory of probability*. Macmillan, London.
Vineis, P. et al. (2004). Tobacco and cancer: recent epidemiological evidence. *Journal of the National Cancer Institute*, 96(2):99–106.
von Mises, R. (1928). *Probability, statistics and truth*. Allen and Unwin, London, second (1957) edition.
Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London.
Weed, D. (2000). Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *International Journal of Epidemiology*, 29:387–390.
Williams, M. and Williamson, J. (2006). Combining argumentation and Bayesian nets for breast cancer prognosis. *Journal of Logic, Language and Information*.
Williamson, J. (2005). *Bayesian nets and causality: philosophical and computational foundations*. Oxford University Press, Oxford.
Williamson, J. (2006a). Dispositional versus epistemic causality. *Minds and Machines*.
Williamson, J. (2006b). Motivating objective Bayesianism: from empirical constraints to objective probabilities. In Harper, W. L. and Wheeler, G. R., editors, *Probability and Inference: Essays in Honour of Henry E. Kyburg Jr.* College Publications, London.
Williamson, J. (2006c). Philosophy of probability: objective Bayesianism and its challenges. In Irvine, A., editor, *Handbook of the philosophy of mathematics*. Elsevier.