

CHAPTER 15

APPLIED EPISTEMOLOGY AND ARGUMENTATION IN EPIDEMIOLOGY

Mark Battersby

A main cause of philosophical disease – a one-sided diet: one nourishes one’s thinking with only one kind of example.

Wittgenstein. *Philosophical Investigations* p.593

1. APPLIED EPISTEMOLOGY AND ARGUMENTATION

This paper is a further development of the concept of “applied epistemology” that I first proposed in a paper in *Informal Logic* (Battersby 1989). After explaining the idea of applied epistemology, this paper focuses primarily on the science of epidemiology and what “applied epistemologists” (né informal logicians) can learn from the epistemological practices used in epidemiology. In the spirit of the Wittgenstein quotation, I invite those who are interested in applied epistemology and are looking for a model of how a “hard” science actually establishes causal claims to look at epidemiology, rather than the traditional paradigm of physics. Epidemiology is a highly successful science and, to some extent, epistemically self-conscious. It is not characterized by over-arching laws à la Newton, nor does it lend itself to the application of the Popperian principle of falsifiability. Because epidemiology is fundamentally a stochastic science, and no experi-

ment is sufficiently conclusive to falsify a claim, falsification is as elusive as proof. Despite that, epidemiology has had enormous success in contributing to both an understanding and an enhancement of human health through the identification of the causes of diseases and to the resultant development of crucial public health recommendations. But first a bit of background on the idea of applied epistemology.

1.1. Why “applied epistemology” and how does it relate to argumentation?

The Stanford Encyclopaedia of Philosophy’s definition of “informal logic” is:

. . . an attempt to develop a logic which can be used to assess, analyse and improve the informal reasoning that occurs in the course of personal exchange, advertising, political debate, legal argument, and in the types of social commentary found in newspapers, television, the World Wide Web and other forms of mass media (Groarke).

I rejected this view of “informal logic” in the earlier paper referred to and argued that the enterprise was better thought of as “applied epistemology,” analogous to applied ethics. The terms “informal logic” tends to anchor¹ the study of arguments in formal logic. Such a nomenclature tempts us to use models of reasoning based on deduction and to potentially miss the actual nature of most reasoning. “Applied epistemology,” on the other hand, focuses the discipline towards the actual practice of how people come to and should come to justified beliefs. On analogy with applied ethics, the study of people’s actual epistemological practices can provide both information and challenges for the theoretician of reasoning.

Applied ethics has created a robust research project and

1. The concept of “anchoring” is used in psychology to describe the tendency of people to be non-rationally influenced by where ever they start their deliberations. For example, in buying real estate, the asking price often influences people’s offers independent of the worth of the property.

stimulated ethical thinking both inside and outside philosophy. Studying and theorizing about the epistemological and argumentative practices of other disciplines may yield comparable insights. There is no reason for applied epistemology (or informal logic) to limit itself to the study of popular arguments as described in the above definition. “Informal” reasoning, argumentation, is the most important reasoning in virtually every discipline. Even those disciplines characterized by a high degree of mathematization (such as epidemiology) still involve non-formal arguments. The only exception may be mathematics itself. Studying how professionals in other fields actually reason (the arguments that they actually make in support of their claims) and how they evaluate claims, provides important information for any theory of applied epistemology — just as studying how medical practitioners make moral decisions informs applied ethics. Philosophers who focus on the norms of informal reasoning and argumentation may well be able to contribute to other disciplines by suggesting ways to improve reasoning and epistemological evaluation in those disciplines. However, applied philosophy is not just about philosophy being “useful,” it is also about learning from the practices of “reflective practitioners.” The place of applied epistemology in relation to epistemology generally can be seen in the following table that sketches my view of the parallels between ethics and epistemology.

Level	Ethics topic examples by level	Epistemology topic examples by level
Meta-ethics/ epistemology	Meaning of “Good”	Meaning of “know”
Normative ethics/ epistemology	Utilitarianism vs. deontology	Rationalism vs. empiricism
Applied ethics/ epistemology	Criteria for morality acceptable euthanasia	Criteria for accepting a causal claim

Applied epistemology also focuses an approach to argumentation on epistemological criteria rather than on rules for nor-

matively correct dialogue and discourse, the approach often favoured in argumentation theory. This paper will illustrate how the analysis of argumentation in epidemiology can contribute to the identification of criteria for justifying causal claims and will also explore in what ways argument analysis can contribute to the improvement of both the criteria and their use in argumentative discourse.

1.2. Epidemiology

What is epidemiology? Below are two typical definitions:

Epidemiology: a branch of medical science that deals with the incidence, distribution, and control of disease in a population. (*Merriam-Webster Online*)

Epidemiology: Epidemiology is the study of the distributions and (causal) determinants of disease in populations. (from the *Dictionary of Epidemiology* 62, (John M. Last, ed., 4th ed. 2001, quoted by Weed 2004)

These are typical definitions, but I believe that a more descriptively accurate definition would be:

The scientific study of human health and illness based primarily on the statistical study of human populations.

This definition allows epidemiology to study everything from the Atkins Diet, the costs and benefits of using estrogen with postmenopausal women to the spread of avian flu and the effectiveness and dangers of Vioxx. Epidemiologists are usually medically and statistically trained researchers.

Epidemiology provides an excellent discipline for the applied epistemologist to study because, despite using rigorous statistical methods, claims to have established correlations and causal relationships must be defended through argument involving a large range of complex considerations. This claim may seem surprising to anyone who has looked at medical research since most research emphasizes statistical

concerns such as whether claims are “statistically significant.” But, in fact, few studies actually meet the random sampling criteria for the application of these statistical methods. Therefore, researchers must argue for the credibility of their results, not merely apply a formula. Justifying a causal claim requires even more arguments than for a correlation. Epidemiologists must argue for any causal claim they make using a variety of relevant considerations. Claims are seldom established by critical experiments or the confirmation of a precise prediction. Rather, they are established by an evaluation of numerous relevant considerations — as they are in many sciences. Establishing a causal claim typically involves making a case (i.e., argument) that appropriate epistemological norms, such as the following, have been satisfied:

- The correlations identified are reliable.
- Confounding factors were appropriately controlled.
- Biological analogies from animal experiments, other lab experiments, and accepted biological theories support the claim.
- Counter-arguments and objections can be dealt with effectively.

2. CAUSALITY IN EPIDEMIOLOGY

2.1. History and the development of criteria

It is informative to study the history of epidemiology from an epistemological perspective. In the 19th and 20th centuries, the field of epidemiology went through a series of fundamental revisions as to how causal claims should be established. Early epidemiologists, such as the famous John Snow, whose work helped prevent cholera epidemics in mid 19th century London, did not have models of the causal mechanism for the spread of disease. Because of this lack, they were

restricted to establishing correlations between exposure and illness. For example, Snow identified a correlation between certain water sources and the incidence of the cholera. Lacking a biological theory, early epidemiologist could only speculate on possible linking causes. Today, epidemiologists utilize not only statistical methods, but also whatever biological models are available to establish causal relationships between causal factors and health outcomes: e.g., broccoli leads to reduced cancer, bacteria lead to ulcers. Claims are established by combining the statistical results of studies and results from laboratory experiments together with the best biological knowledge.² Epidemiologists study not only causes of illness but also putative cures. The studies that confirmed the viability of the polio vaccine are one famous example of epidemiology in service of preventative medicine.

My focus on epidemiology as a paradigmatic science is not without a somewhat ironic precedent in analytic philosophy. Carl Hempel, in his classic *Philosophy of Natural Science* (1966), used an account of the effort of an early epidemiologists, Ignatz Semmelweis, to introduce scientific reasoning. Hempel describes at some length Semmelweis' efforts to discover the cause of a higher incidence of puerperal fever in one of the two maternity wards in his hospital. As many will recall, Hempel uses Semmelweis' story to illustrate how science often proceeds by trial and error and the elimination of competing hypotheses. Despite beginning with this story, Hempel goes on to theorize about causal explanations largely with reference to reasoning in physics not medical research.

As Hempel records, Semmelweis theorized that the cause of higher mortality from so-called "puerperal fever" in one of the two maternity wards was due to "cadaverous matter" on the

2. Actually, there is still a debate within epidemiological circles over whether to take a "black box" approach and just crunch number, or to incorporate biological theories. This approach is often embodied in the use of terms like "risk factor" which avoids having to make a causal claim

hands of medical students emerging from the nearby autopsy room before examining the pregnant women in that ward. By having the students wash their hands, Semmelweis was able to reduce the level of mortality in the higher mortality ward to a rate comparable to that in the other. Regrettably, there still was a 3% mortality rate in both wards which underlies the complexity of epidemiological causal reasoning: cadaverous matter was neither necessary (3% were infected anyway) nor sufficient for the illness (the rate in the ward with higher mortality was 9%). And as we all know, it was not only matter derived from cadavers that caused the illness. Semmelweis himself later theorized that it was “putrid” matter because he realized that the illness was being transmitted from the sick, not just the dead.

One of the theories that Semmelweis rejected before his discovery was the theory that puerperal fever was caused by “cosmic telluric changes.” This type of causal theory was a common place in early medicine—ascribing many illnesses to a general miasma that just affected some people.

In the late 19th century, as the germ theory of illness gained acceptance, this miasma approach to aetiology was rejected by the renowned German pathologist, Jakob Henle and his student Robert Koch, who articulated the following rigorous criteria for a causal claim in medicine:

- The agent should be present in every case of the disease under appropriate circumstances.
- The agent should not be present in any other disease as a fortuitous and nonpathogenic agent.
- The agent must be isolated from the body of the diseased individual in pure culture, and it should induce disease anew in a susceptible animal (Pai 2005).

Helpful and rigorous as these criteria were, they later required extensive revision as the study of disease moved

from a focus on pathogens to a focus on a complex of factors. The presupposition of one disease/one pathogen just did not fit emerging facts about such illnesses as cancer. For example, the research into smoking that was done in the early 50s revealed a strong association between smoking and lung cancer, but also, a strong association with coronary artery disease. Critics of the day argued, using the Henle-Koch criteria, that this showed that smoking could not be the true cause of lung cancer (Stolley, p.65). Rather than accept this criticism, researchers began to develop alternative *criteria* that would form the basis for establishing causal claims about diseases.

The 1964 Surgeon General's Committee on Smoking and Health developed explicit criteria to determine whether smoking caused the diseases under review because of the public scrutiny to which their study would be subjected. The list included (with my comments):

- *Consistency of findings.* Conflict in evidence mitigates against a causal claim.
- *Strength of association.* The dramatically high relative risk of lung cancer among smokers was a crucial basis for the causal claim.
- *Specificity.* A bit of a left over from previous criteria, though the committee points out that smokers only have higher mortality in a few other diseases
- *Temporality.* Cause must occur before effect
- *Biological coherence.* Under which they included biological mechanisms and fit with existing understanding, biological models and animal experiments.
- *Dose-response.* More tobacco use correlated with a higher lung cancer rate.
- *Exclusion of alternate explanations.* Such as bias but also

competing explanations such as 3rd causes (e.g., genetic tendency to both smoke and get cancer).

A year later, Bradford Hill, a leading biostatistician, articulated the following slightly more complex set of considerations (he called them “viewpoints”). Strangely, he left out consideration of the exclusion of alternative explanations, which is, of course, crucial to making a “causal case.” His approach ignores, as I will argue below, that making argument for a causal claim is really best seen as “argument to the best explanation.” The justification for rejecting competing explanations is central to such an argument. So crucial is the rejection of competing explanations that other theorists include it under “Hill’s Criteria” (Arbuzzi 2005).

- *Strength.*
- *Consistency.*
- *Specificity.* Still left over from Henle-Koch but often reinterpreted as high strength of association
- *Temporality.* A cause must precede an effect in time.
- *Biological gradient.* Dose-response relationship.
- *Plausibility.* The idea of causation must be biologically plausible.
- *Coherence.* The idea of causation must accord with other observations.
- *Experimental evidence.* Supporting data from human or animals experiments, such as lung cancer in animals exposed to cigarette smoke, helps establish a causal relationship.
- *Analogy.* For example, if thalidomide can cause birth defects, perhaps other drugs taken during pregnancy can also cause birth defects. Analogy can be helpful, although the help seems limited since anybody with a little creativity can probably dream up an analogy.

Hill's criteria are neither necessary nor sufficient for ascribing causality. They are analogous to a set of considerations that one might suggest for moral decision making such as Ross's famous list of *prima facie* duties³ or any procedure of moral reflection that invites one to consider a list of crucial considerations such as: 1. the rights of individuals affected, 2. the relevant obligations, both general and specific (e.g., occupational), 3. the consequences to all parties affected, etc.

As in ethical reflection, different researchers emphasize different criteria at different times. This could be a bad sign if it revealed inconsistency or bias. As with most disciplines, epidemiology is not characterized by a consistent epistemological self-consciousness. While frequent mention is made of the "Hill Criteria," researchers tend to refer only to a convenient sub-set. It is an open question (discussed briefly below), whether a precise list of weighted criteria could be developed. Nevertheless, the example below, on the efficacy of prayer, suggests that a more reliable use of criteria could eliminate at least egregious examples of implausible claims.

2.2. The need for criteria

The following is an entertaining demonstration of the need for the application of epistemological criteria and for understanding that a claim needs argument, not just methodologically sound statistics. This study appears to violate almost every criterion for establishing a causal claim and yet was published in the *British Journal of Medicine* in 2001. I believe it was published because of the respect accorded by editors to the norm of statistical significance. The criterion of statistical significance is simply a statistical convention for determining

3. Ross 1930. Ross' list: Fidelity: the duty to keep promises, Reparation: the duty to compensate others when we harm them, Gratitude: the duty to thank those who help us, Justice: the duty to recognize merit, Beneficence: the duty to improve the conditions of others, Self-improvement: the duty to improve our virtue and intelligence, Nonmaleficence: the duty to not injure others.

that an apparent correlation is probably not due to chance. Regrettably, statistical significance often serves as both a necessary and sufficient condition for publication.

The study by an Israeli researcher, Leonard Leibovici, was entitled “Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: a randomized controlled trial.”

Abstract

Objective: To determine whether remote, retroactive intercessory **prayer**, said for a group of patients with a bloodstream infection, has an effect on outcomes.

Design: Double blind, parallel group, randomised controlled trial of a retroactive intervention.

Setting: University hospital.

Subjects: All 3393 adult patients whose bloodstream infection was detected at the hospital in 1990-6.

Intervention: In July 2000, patients were randomised to a control group and an intervention group. A remote, retroactive intercessory **prayer** was said for the well-being and full recovery of the intervention group.

Main outcome measures: Mortality in hospital, length of stay in hospital, and duration of fever.

Results: Mortality was 28.1% (475/1691) in the intervention group and 30.2% (514/1702) in the control group (P for difference=0.4) [*i.e. this result does not meet the typical criteria for statistical significance of <.05*]. Length of stay in hospital and duration of fever were significantly [*i.e., statistically significant*] shorter in the intervention group than in the control group (P=0.01 and P=0.04, respectively).

Conclusions: Remote, retroactive intercessory prayer said for a group is associated with a shorter stay in hospital and shorter duration of fever in patients with a bloodstream infection and should be considered for use in clinical practice.

Unsurprisingly this study produced a stream of protest letters, but many letter writers failed to point out the conflict

with the temporality condition. Only one writer identified the obvious alternative explanation that it was simply a statistical fluke. As all statisticians know, what the claim of *statistical significance* means in this context is that there was only a 1/100 or 4/100 chance that the results would occur by chance. Rare, but hardly out of the question, and a lot more credible explanation than the causal efficacy of *retroactive* prayer.

2.3. The tempting illusion of statistical precision

It is the sign of an educated man that in every subject he studies, he seeks only that degree of precision which the nature of the subject permits (e.g., it is absurd to expect logic from a public speaker or probabilities from a mathematician) (Aristotle, *Nicomachean Ethics*, 1094b23-28).

In view of the somewhat unreliable way in which the criteria are used, various efforts have been made to articulate a tighter set of criteria. Predictably, there is also increased interest in finding more algorithmic approaches.

While no doubt something will be learned by such a formalization project, the effort to formalize the inference from evidence to causality seems unlikely to succeed. There are just too many factors that are difficult to quantify to establish a realistic mathematical measure. There is also a danger that the use of mathematics will create an appearance of precision that is misleading. Even the current use of statistical inference in epidemiological research is often misleading. For example, almost no studies meet the condition of random sampling which provides the mathematical basis for applying the formulae. The so-called “case controlled studies” which play an important role in epidemiological research consist of matching a group of people who have an illness with a comparable group of people who don’t have the illness and then looking for factors that are more prevalent among the ill than among the controls.

Obviously, the choice of comparable controls can have a

great effect on the utility of the comparison. Yet there are not and cannot be mathematical standards for selecting the controls. The controls are selected on assumptions about what aspects of an individual are crucial for identifying relevant similarity. The obvious factor of age is almost always taken into account, but even gender and race are frequently ignored. And what else is missing?

To see how this works in practice, take the case of early studies into the smoking/lung cancer link. In the early 1950s, two retrospective studies of approximately 600 to 700 cases of lung cancer were done that compared the history of smoking among lung cancer victims and “control” groups made up of other hospital patients of “similar” characteristics who did not have lung cancer. The samples of subjects used in this approach are known as “samples of convenience.” Both of these early studies found a slightly higher rate of smoking among the cancer victims than the control group, but the differences between the rates were not great enough to be statistically significant, i.e., the researchers could not be 95% confident that the differences in the rate of smoking between the groups was not due to chance. Researchers still believed there was a relationship between smoking and lung cancer, although their study had failed to “statistically” demonstrate it. Why had the study failed to demonstrate what is, in fact, a strong correlation? With the advantage of hindsight, we can clearly see the problem. None of the patients in the “control group” had lung cancer, but many of them had illnesses to which we now know smoking contributes (such as heart disease). The control group was not representative of the non-lung cancer population. The controls had a larger percentage of smokers than in the non-lung cancer population of comparable age. The unrepresentative percentage of smokers in the control group obscured the actually dramatic difference in the rate of lung cancer between smokers and non-smokers (Cornfield 1959, p.182).

This is not just a problem in scientific research. While it is widely believed that the ideal sample for polls is a “representative” sample of the population, pollsters have learned the unreliability of such samples. The famous pollster, George Gallop, initially gained great renown in the 1940s when he used *representative* sampling to more or less correctly predict the re-election of Roosevelt. His poll was based on the sampling of some 8000 people, in contrast to the *Literary Digest* poll which surveyed millions and made the wrong prediction. Nonetheless, when Gallop used the same technique for the subsequent Truman election, he predicted the wrong victor and his prediction was badly off. Subsequently he went to random sampling *not* representative sampling, recognizing that it is not possible to reliably identify the factors that make for a representative sample. Gallop’s lesson has not been reflected in most scientific research simply because such random selection techniques usually cannot be used in this research. Participants in studies are necessarily volunteers who were not randomly selected and many diseases have too low an incidence to be effectively studied using random selection. My point is not to deride the research, but to re-emphasize that judgment and argument (not probability theory) must be used to support the claim that the samples and control groups that were studied provide a reasonable basis for the correlational and causal claims being made.

2.4. Argumentation in epidemiology

As argued above, statistical inference is often not adequate for establishing correlations in most studies. It is never adequate for establishing causal claims. Correlations are necessary but not sufficient for a causal claim. Epidemiologists must, therefore, use informal arguments to make their case for a causal claim. Basically, what epidemiologists do is argue that their claim is the best explanation. While the status of “inference to the best explanation” as the best account of scientific

reasoning remains controversial in philosophy, it seems clear that the argumentative process in epidemiology is best characterized in this way. The primary objection of philosophers to “inference to the best explanation” account of scientific reasoning is that the notion of “best explanation” is vague and/or circular. But if we take an applied epistemological approach to analyzing the work of epidemiologists, we can see how they use the criteria discussed above to substantiate their positive claims and reject counter theories.

One of the most famous and effective examples of what I wish to call “*argument for the best explanation*” was made in 1959 by Jerome Cornfield and others arguing the case that smoking is the primary cause of lung cancer. This article is widely considered to have established the case for smoking as a cause of lung cancer and led to public policy efforts such as the Surgeon General’s Report cited above.

In his summary, Cornfield both argues for his claim and rejects alternative hypotheses:

The magnitude of the excess lung cancer risk among cigarette smokers is so great that the result cannot be interpreted as arising from an indirect association of cigarette smoking with some other agent... The consistency of all the epidemiological and experimental evidence also supports the conclusion of a causal relationship ...while there are serious inconsistencies in reconciling the evidence with other hypotheses which have been advanced (Cornfield 1959, p.173).

In his article, Cornfield first reviews the existing literature in support of the causal claim, and then devotes most of the paper to responding to criticisms of the studies. He divides the responses into 5 major topics:

- population data
- retrospective and prospective studies
- studies on pathogenesis

- other laboratory investigation
- interpretation

In the first section, he replies to the objection that the significant difference in the rate of lung cancer among men and women is grounds for discarding the causal hypothesis. He points out that the data shows that men have been smoking for significantly longer than women, especially in the over 55 age group, which is the demographic that mainly experiences lung cancer. In addition, he notes that the rate of lung cancer among both male and female non-smokers is similar.

In a section on criticisms of retrospective studies, Cornfield argues: “. . . for the most part, the specific points of criticism apply only to some of the studies and not to others” (p.181). He argues for the overall convergence of the research despite specific problems with any particular study.

In another section, Cornfield replies to the objection that experiments involving rats exposed to smoke have failed to induce lung cancer, as being “. . . true at the time of this report, although it can be questioned whether any animal received as large a dose of cigarette smoke through indirect exposure as a human being does by voluntary deep inhalation.” He had earlier noted the difference in rates of lung cancer among inhalers and those that did not inhale.

Cornfield acknowledges that nothing short of randomized trials could provide a clear cut answer to what he calls the “constitutional hypothesis,” the idea that some people are prone genetically to both smoke and get lung cancer. Nevertheless, he argues this hypothesis is inconsistent with the following observations:

1. changes in the lung cancer mortality over the last half-century,
 2. carcinogenicity of tobacco tars for experimental animals,
 3. effect of pipe smoking on larynx cancer but not lung cancer,
 4. reduced lung cancer among discontinued smokers.
- No one of these considerations is perhaps sufficient by itself to discount

the constitutional hypothesis, ad hoc modifications of which can accommodate each additional piece of evidence. A point is reached, however, when a continuously modified hypothesis becomes difficult to entertain seriously (Cornfield 1959, p.191).

Lastly, Cornfield replies to the well-known question of why many smokers never get lung cancer: “We have no answer to this question. But neither can we say why most of the Lubeck babies who were exposed to massive doses of virulent tubercle bacilli failed to develop tuberculosis [*note the argument by analogy*]. This is not a reason, however, for doubting the causal role of the bacilli in the development of the disease” (p.197).

The foregoing are only a sample of the arguments that fill the 30-page article. But as can be clearly seen, they involve a wide variety of informally presented appeals to science and common sense. In fact, the only statistical part of his response is placed in an appendix. Cornfield’s paper was published before the Surgeon General and Bradford Hill published their epistemological reflections. Nonetheless, a detailed study of his arguments reveals that he employs the notions of:

- *Strength*. He cites the high relative risk of lung cancer for smokers.
- *Consistency*. As mentioned, almost all studies point in the same direction.
- *Specificity*. Here the issue is to confirm that the relation is not actually the result of other factors where smoking is just a token for these factors. For example, smokers have a higher mortality rate from all causes, not just lung cancer, which would suggest that something else could be at work in the lung cancer – smoking association. But in response, Cornfield points out that these correlations are weak compared to that of smoking and lung cancer.
- *Temporality*. He emphasized the lag time between exposure and cancer to explain some apparent anomalies.

- *Biological gradient.* Heavier smokers get lung cancer at a higher rate.
- *Plausibility.* He speculates on possible causal models while admitting this is a weakness in the argument.
- *Coherence.* The lung cancer result fits with the fact of higher levels of upper respiratory cancer in pipe smokers who do not inhale.
- *Experimental evidence.* Rats painted with tars had high rate of skin cancer.
- *Analogy.* Cited above, re-exposure not necessarily producing disease.
- *Exclusion of alternative explanations.* Argument against the genetic theory above.

Notice that no explicit weighting is given. He simply marshals the overall evidence, replies to critics, and shows that the weight of evidence supports the causal hypothesis.⁴

Experimental	Inferential
Analysis of a single study	Integration of multiple studies
Randomization essential	No "crucial experimentation"
Specificity of association	Strength of association

While Passcandola's contrasts are not quite parallel, the table provides a useful brief summary of the issue seen from inside the discipline. Historically, the experimentalist lost the smoking/lung cancer debate, though introductory books on experimental method and statistics (largely written by statisticians) still tend to emphasize the former approach (cf.

4. It should be admitted that my view of the epistemology of epidemiology is not universal in that discipline. In an informative overview of the history of the smoking and lung cancer debate, Mark Passcandola (June 2004) identifies two approaches which he calls the experimental and the inferential. He contrasts them as follows:

the generally excellent introduction statistics book by Jessica Utts, 2005).

3. APPLYING EPIDEMIOLOGICAL CAUSAL CRITERIA TO OTHER DISCIPLINES

The criteria used by epidemiologists to make their argument that their causal claim is the best explanation may also be used in other disciplines. For example, the debate over the causal effects of pornography continues although currently at a much lower key than in the late 20th century. This issue, like many of those in epidemiology (such as the causal effect of passive smoke) has profound public policy implications. Those who argue for the negative effects of pornography have a fairly strong burden of proof as they are up against the strong presumption in favour of free speech.

A recent review of the research by a student of mine, Lindsay Johnson, found that such strong evidence was difficult to find and that, in fact, there was some powerful counter-evidence that suggested another, far more significant causal factor. In her study, she cited work by Dodson which makes the following claims (I have indicated in italics the various causal considerations that are implicitly appealed to):

Studies on violent pornography are inconsistent. Some find it increases aggression in the lab; some find it does not. Research also finds that aggression will be increased by anything that agitates a subject (that raises heart rate, adrenaline flow, etc.), not only violent movies but riding exercise bicycles. Agitation will boost whatever follows it, aggression or generosity. (*lack of specificity, alternative explanations*)

Dr. Suzanne Ageton, measuring violence out of the lab, found that membership in a delinquent peer group accounted for 3/4 of sexual aggression. (*alternative explanation*)

Studies in the U.S., Europe, and Asia find no link between the availability of sexual material and sex crimes. The only factor linked to rape rate is the number of young men living in a given area. When pornography became widely available in Europe, sexually violent crimes decreased or remained the same. Japan,

with far more violent pornography than the U.S., has 2.4 rapes per 100,000 people compared with the U.S. 34.5 per 100,000. (*no evidence of “dose” relationship*)

Since the difficulties of establishing causal claims are probably even more complex in the social sciences than in epidemiology, I would suggest the social sciences could also benefit from making the case for their claims using “argument to the best explanation” and making appropriate use of epidemiological criteria when doing so. Neither of the two famous efforts by the United States government to address the causal effects of pornography displayed the kind of epistemological self-consciousness shown in the Surgeon General’s Report on Smoking referred to above.

4. HOW MIGHT “APPLIED EPISTEMOLOGISTS” CONTRIBUTE TO WORK IN EPIDEMIOLOGY? JUDGMENT AND THE PROBLEM OF BIAS IN EPIDEMIOLOGY

Cornfield’s paper illustrates that judgment and argument play a central role in the assessment of causal claims. Unfortunately, judgment and argument provide considerable opportunity for bias. The natural sciences, because of their emphasis on “letting the data speak for themselves” have been largely able to avoid the kind of epistemologically undermining influence that bias plays in say political “science” or economics. Nonetheless, as the historic debate about the effects of smoking and recent pharmaceutical testing scandals illustrate, bias can be a crucial factor in epidemiological work. Fair-mindedness and a careful respect for both the significance and difficulties of any research are important in any discipline, but are crucial in one in which arguments and “judgment calls” are central.

Such observations have implications not only for the administration of scientific funding, but also for the adjudication of scientific results. What evaluative weight, for example, should be given to the fact that research was funded by a man-

ufacturer? How can we make appropriate use of a researcher's statements of conflict of interest without slipping into the *ad hominem* fallacy?

The debate over passive smoking, or more technically, Environmental Tobacco Smoke (ETS), illustrates many of these problems. The studies in this area exhibit much more conflict and, not surprisingly, a much weaker association between smoke exposure and lung cancer incidence. The commonly cited risk factor of 1.2 (an average of many studies') means that people who are exposed to ETS have an approximately 20% higher risk of getting lung cancer than those who are not exposed. This is in contrast to the relative risk of smokers which is between 6-16 times the risk of non-smokers (depending on amount smoked). An additional problem with ETS research is determining the amount of exposure.

Two recent studies related to ETS illustrate both the difficulties involved in the research and the problem of evaluating the appearance of bias without descending into the *ad hominem* fallacy.

An article by **James E. Enstrom** and **Geoffrey C. Kabat** published in the *British Journal of Medicine* (Enstrom and Kabat 2003) caused a storm of protest when it published the following results from a prospective study of 120,000 Californians: "For participants followed from 1960 until 1998 the age adjusted relative risk (95% confidence interval) for never smokers married to ever smokers compared with never smokers married to never smokers was 0.94 (0.85 to 1.05)." That is, they failed to find a correlation between spousal exposure and increased lung cancer rate. Enstrom and Kabat concluded: "The results do not support a causal relation between environmental tobacco smoke and tobacco related mortality, although they do not rule out a small effect."

The authors admitted in their statement of interests that:

In recent years JEE (**James E. Enstrom**) has received funds orig-

inating from the tobacco industry for his tobacco related epidemiological research because it has been impossible for him to obtain equivalent funds from other sources. GCK (**Geoffrey C. Kabat**) never received funds originating from the tobacco industry until last year, when he conducted an epidemiological review for a law firm which has several tobacco companies as clients. He has served as a consultant to the University of California at Los Angeles for this paper. JEE and GCK have no other competing interests. They are both lifelong non-smokers whose primary interest is an accurate determination of the health effects of tobacco.

Much was made of the authors' tobacco industry association in the subsequent firestorm of objections to the paper.

So virulent was the attack (which also involved arguments that *BJM* should not have published the paper because of the comfort it would give to the tobacco lobby) that the editor of *BJM* felt the need to respond:

Firstly, we've considered again whether we should have a blanket policy of refusing to publish research funded by the tobacco industry. We've twice considered this question in the *BMJ* and twice decided against. The *BMJ* is passionately antitobacco, but we are also passionately prodebate and proscience. A ban would be antiscience.

Secondly, we are not in the "truth" business. Scientific truths are all provisional. Most of science falls away as new paradigms emerge. This doesn't mean that we are in the "lies" business, but we are in the "debate" business.

Thirdly, with research papers we first ask if we are interested in the question. We must be interested in whether passive smoking kills, and the question has not been definitively answered. It's a hard question, and our methods are inadequate.

We then peer review the study, but we are well aware of the extreme deficiencies of peer review. Of course the study we published has flaws—all papers do—but it also has considerable strengths: long follow up, large sample size, and more complete follow up than many such studies. It's too easy to dismiss studies like this as "fatally flawed," with the implication that the study means nothing.

Fourthly, I found it disturbing that so many people and orga-

nizations referred to the flaws in the study without specifying what they were. Indeed, this debate was much more remarkable for its passion than its precision. **Richard Smith**, *editor*

As Smith's remarks indicate, many of the criticisms suffered from the *circumstantial ad hominem fallacy*. In fact, one of the authors in responding to the accusations argued: "Scientists, and particularly epidemiologists, who deal with the criteria for judging causality, should be wary of imputing motives based on the *flawed logic of guilt by association*."

Whatever the flaws in the study, it seems clear that the suspicion of bias and the role of tobacco funding played a crucial role in the debate. Were the critics who objected to the authors' funding all guilty of the *ad hominem* fallacy? What weight should be given to the authors' funding sources? Interestingly, there is "epidemiological" evidence that some weight should be given. A 1998 article also in the *British Journal of Medicine*, by Barnes and Bero entitled "**Why Review Articles on the Health Effects of Passive Smoking Reach Different Conclusions**" argued that bias was definitely at work in passive smoking research.

Abstract

The authors reviewed review articles on the topic of ETC and found that:

Data Synthesis. A total of 106 reviews were identified. Overall, 37% (39/106) of reviews concluded that passive smoking is not harmful to health; 74% (29/39) of these were written by authors with tobacco industry affiliations. In multiple logistic regression analyses controlling for article quality, peer review status, article topic, and year of publication, the only factor associated with concluding that passive smoking is not harmful was whether an author was affiliated with the tobacco industry (odds ratio, 88.4; 95% confidence interval, 16.4-476.5; $P < .001$).

Conclusions. The conclusions of review articles are strongly associated with the affiliations of their authors. Authors of

review articles should disclose potential financial conflicts of interest, and readers of review articles should consider authors' affiliations when deciding how to judge an article's conclusions (**Barnes and Bero 1998**).

While the numbers in the abstract are a bit incomprehensible, there does seem to be a strong *prima facie* case that bias is at work in this area of research. But we should be careful. The claim of funding bias is that the funding is causally related to the judgment in the study. But all that the evidence establishes is that there is a correlation. We must be careful about the inference to causality, in particular the application of the criteria of *temporality*. Funding support may follow research that happens to support the position desired by willing funder rather than researchers being paid to do studies that support the funder's point of view. This appears, for example, to be what happened in the passive smoking article cited above.

How should readers "consider the affiliations of the author"? As the comments by the editor of *British Journal of Medicine* indicate, what to do about corporate funding in science is a huge question. Disclosure of financial interests certainly seems essential, but clearly such disclosure may result in the fallacious dismissal of legitimate research. If you believe that any use of *ad hominem* observations in an argumentative context is fallacious (and irrelevant), then you would not even require that authors cite their funding sources. The reason that *ad hominem* remarks are often fallacious, as the BJM editor notes, is that they tempt people to facile dismissal without looking at the details of the study. On the other hand, the problem with ignoring information about the authors' funding support (or even publication record) is that this is clearly information that can help contextualize (though not refute) an author's argument. I believe that most informal logicians would support the BJM editor and the article's authors in discouraging people from solely basing their judgments of a study on the basis of an author's funding sources, but would

also support a policy of requiring authors to acknowledge their funding sources. To understand the breadth of this issue, it should be noted that all testing of new drugs is funded by pharmaceutical companies.

5. APPLICATION: EXPLORING THE RELATIONSHIP BETWEEN ARGUMENTATION, APPLIED EPISTEMOLOGY AND EPIDEMIOLOGY

5.1. Applying critical thinking to reading medical research

In *Evidence Based Practice: Logic and Critical Thinking in Medicine* (Jenicek and Hitchcock 2005), the authors do a masterful job of describing a critical thinking approach to epidemiological reasoning – what I would call an excellent example of *applied epistemology*. The authors use work in critical thinking and epidemiology to lead the student through the appropriate reasoning processes for argumentation in medicine and for the assessment of causal claims. They provide a list of considerations that articulate the criteria for justifying causal claims in epidemiology, basing their list on a number of contemporary textbooks.

Assumptions (prerequisites, before any causal criteria apply)

- Exclusion of the play of chance
- Consistency of results with prediction
- Even observational studies respect as much as possible the same logic and similar precautions as used in experimental research
- Studies are based on clinimetrically valid data
- Data are subject to unbiased observations, comparisons, and analysis
- Uncontrollable and uninterpretable factors are ideally absent from the study

Criteria of causation

Major:

- Temporality (“cart behind the horse”)
- Strength (relative risk, odds ratio, hazard ratio)
- Specificity (exclusivity or predominance of an observation)
- Manifestational (“unique” pattern of clinical spectrum and gradient as presumed consequence of exposure)
- Causal (attributable risk, etiological fraction, attributable risk percent, attributable hazard, proportional hazard)
- Biological gradient (more exposure = stronger association)
- Consistency (assessment of homogeneity of findings across studies, settings, time, place, and people)
- Biological plausibility (explanation of the nature of association)

Conditional:

- Coherence with prevalent knowledge
- Analogy

Reference:

- Experimental proof (preventability, curability)
- Clinical trial, other kind of controlled experiment or “cessation study”

Confirmation:

- Systematic review and meta-analysis of evidence

(Jenicek and Hitchcock 2005, 155)

Their list differs from the historical lists cited above, but this should not be surprising. The development and establish-

ment of the criteria is an ongoing example of applied epistemological reflection at work in epidemiology. Jenicek and Hitchcock distinguish between assessment of the data for establishing a correlation (rightly calling these “prerequisites” for applying causal criteria) and criteria for the inference to a causal claim. Unfortunately, from my perspective, they leave out a key basis for a causal claim: the rejection of competing explanations. A further discussion of the criteria and how one might weight them is an issue for another paper (continuing the research project of applied epistemology).⁵

5.2. The symbiotic relationship between informal logic and the epistemological reflections of epidemiologists

To see some of the mutual benefits of looking at the considerations for causal claims identified by epidemiologists and the work of informal logicians, we might compare the Surgeon General’s and Hill’s list to the very credible list of questions that Walton (1989, p.230) uses to evaluate a causal claim. I have changed the order of the various lists to facilitate comparison.

5. The merits of Jenicek and Hitchcock’s work notwithstanding, I do wish to voice a reservation about the authors’ choice of the Toulmin model of argument. This model, with its emphasis on a single warrant between evidence and conclusion does not appear to provide a normatively correct model of the way diverse consideration must be brought to bear when making a judgment of causality. For example, their figure 5-2 (Jenicek and Hitchcock 2005, 165), which is an example of how the authors attempt to use the model, seems to illustrate the limitations of trying to impose the model rather than illuminating how actual arguments should be represented and evaluated.

Surgeon General	Hill	Walton
Consistency of findings	Consistency	Is there a positive correlation between A and B? Are there a significant number of instances of the positive correlation between A and B?
Strength of association	Strength	
Specificity	Specificity	
Temporality	Temporality	Is there good evidence that the causal relationship goes from A to B, and not just from B to A?
Dose-response	Biological gradient	
Biological coherence. <i>biological mechanisms and fit with existing understanding, biological models and animal experiments</i>	Plausibility. <i>The idea of causation must be biologically plausible</i> Coherence. <i>The idea of causation must accord with other observations.</i> Experimental evidence.	
Exclusion of alternate explanations	Analogy	Can it be ruled out that the correlation between A and B is accounted for by some third factor (a common cause) that causes both A and B? If there are intervening variables, can it be shown that the causal relationship between A and B is indirect (mediated through other causes)? Can it be shown that the increase or change in B is not solely due to the way B is defined, the way entities are classified as belonging to the class of Bs, or changing standards, over time, of the way Bs are defined or classified?

If the correlation fails to hold outside a certain range of causes, then can the limits of this range be clearly indicated?

Walton's list is more exhaustive than those found in many in critical thinking textbooks and contains important considerations lacking in Hill's and the Surgeon General's list. Nonetheless, his list omits the importance of the strength of a correlation and ignores the role of explanatory models (biological or others), and the "dose" relationship. On the other hand, his list and the Surgeon General's include the exclusion of alternative explanations.

This is not the place for me to attempt to propose an ideal list, but some comments are, perhaps, apt. A clear distinction needs to be made (as Jenicek does) between criteria for a well-established correlation and criteria for a causal claim. The role of models as explanations (consider the "greenhouse model," for example) needs to be given crucial place in making a strong causal claim, even though epidemiological results often precede detailed biological understanding (see Cornfield). The "juridical" nature of causal claims (we often seek causes in order to assign blame or identify where to intervene) also needs addressing—which may bring in ethical considerations. Ethical considerations will certainly come into play when epidemiologists make recommendations on public policy. The criteria for "announcing" causal claims (while not the same as those for making the claim *simpliciter*) must be epistemically justified while also being related (*à la* Cornfield) to the public policy significance of the finding. The historical context of the debate and issues of onus also need to be addressed. Some of the other criteria referenced in the literature on inference to the best explanation (e.g., simplicity, consilience, etc.) should also be considered. The task is far from easy but it seems clear

that both applied epistemologists and epidemiologists could benefit from sharing this task.

6. SUMMARY

My general goal in this paper was to encourage informal logicians and others interested in applied epistemology to look at epidemiology as a paradigmatic science crucially dependent on argumentation. My two specific goals in this paper were: 1. to give an example of applied epistemology by looking at causal argumentation and justification in epidemiology, and 2. to show that there could be a symbiotic relationship between epidemiology and work in various applied reasoning disciplines such as argumentation, informal logic, philosophy of science and “applied epistemology.”

Epidemiologists are an important example of disciplinary practitioners who develop and apply epistemological criteria. I have argued that epidemiologists would benefit from seeing the justification of a causal claim as making an “argument for the best explanation” which involves not only commonly-used criteria for justifying a causal claim, but also consideration of arguments against alternative explanations. The need for application of some obvious criteria beyond statistical significance was illustrated by the example of the supposed effects of retroactive prayer, and the application of the argument for the best explanation was illustrated by the 1959 paper of Jerome Cornfield on the causal relationship between smoking and lung cancer. I also gave an illustration of how causal criteria used in epidemiology might well be useful in other stochastic sciences such as sociology and psychology.

Of additional interest to informal logicians and argumentation theorist are the dialogic problems that appear periodically in epidemiological discussions around controversial issues such as the effects of passive smoking. The common of use of the *ad hominem* fallacy in these debates represents a shared concern for both informal logicians and epidemiolo-

gists. The appropriate assessment of bias and its relationship to argument evaluation is a topic on which informal logicians should be able to make significant contributions once they take into account the complex role that funding plays in such sciences as epidemiology.

Epidemiology is a rich source of examples for all applied philosophy, but especially applied epistemology. My hope is that this paper will help encourage others to expand their intellectual interests beyond a “one-sided diet” of examples from newspaper editorials or deductivist sciences such as physics.

REFERENCES

- Abruzzi, W. S. 2005. *Hills Criteria of Causation*. Retrieved from http://www.drabruzzo.com/hills_criteria_of_causation.htm April 21.
- Barnes, D.E. and L.A. Bero. 1998. “Why Review Articles on the Health Effects of Passive Smoking Reach Different Conclusions.” *JAMA* 279:1566-1570.
- Battersby, M. 1989. “Critical Thinking as Applied Epistemology: Relocating Critical Thinking in the Philosophical Landscape.” *Informal Logic* 11, 2: 91 – 100.
- Cornfield, J., *et al.* 1959. “Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions.” *Journal of the National Cancer Institute* 22: 173-203.
- Enstrom, J.E. and G.C. Kabat. 2003. “Environmental Tobacco Smoke and Tobacco Related Mortality in a Prospective Study of Californians 1960-98.” *BMJ* 326:1057 (17 May), doi:10.1136/bmj.326.7398.1057)
- Frumkin, H. 2005. *Causation in Medicine*. Retrieved from <http://www.aoc.org/CEEM/methods/emory2.html> April 21.
- Goodman, K.J. & C.V. Phillips. 2005. *Hill Criteria of Causation* (Draft Entry, Encyclopedia of Behavioral Statistics)

- retrieved from http://www.cphps.org/papers/goodman-phillips_abhill-encybehavstat_mar04.pdf April 21.
- Groarke, L. 2003. "Informal Logic." *The Stanford Encyclopedia of Philosophy (Winter 2003 Edition)*, edited by Edward N. Zalta. Retrieved from <http://plato.stanford.edu/archives/win2003/entries/logic-informal/>
- Hempel, C. 1966. *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice Hall.
- Jenicek, M. and D. Hitchcock. 2005. *Evidence Based Practice: Logic and Critical Thinking in Medicine*. American Medical Association: AMA Press.
- Leibovici, L. 2001. "Effects of Remote, Retroactive Intercessory Prayer on Outcomes in Patients with Bloodstream Infection: Randomized Controlled Trial." *BMJ* 323:1450-1451.
- Pai, M. 2005. "Associations and Causations." Retrieved from <http://www.sunmed.org/caus.html> April 21.
- Parascandola M. and D. L. Weed. 2001. "Causation in Epidemiology." *Journal Epidemiol. Community Health* 55: 905-912.
- Parascandola M. 2004. "Two Approaches to Etiology: The Debate Over Smoking and Lung Cancer in the 1950s." *Endeavour* 28, 2:81-86.
- Ross, W.D. 1930. *The Right and The Good*. Oxford: Oxford University Press.
- Stolley, P. and T. Lasky. 1995. *Investigating Disease Patterns: The Science of Epidemiology*. New York: Scientific American Library.
- Utts, J.M. 2005. *Seeing Through Statistics*, 3rd edition. Davis, CA: University of California Press.
- Walton, D. 1989. *Informal Logic: A Handbook of Critical Argumentation*. Cambridge: Cambridge University Press.
- Weed, D.L., & L.S. Gorelic. 1996. "The Practice of Causal Inference in Cancer Epidemiology." *Cancer Epidemiol Biomarkers Prev* 5: 303-311.

Weed, D.L. 2004. "Causation: An Epidemiologic Perspective" (In Five Parts). *Journal of Law and Policy*: 43-53.