

2nd North American Congress of Epidemiology, Seattle, WA, June 23, 2006

USING EPIDEMIOLOGIC EVIDENCE TO ADVANCE HEALTH: DEALING WITH CRITICS AND CRITICISMS. Jonathan M. Samet (Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205) (http://aje.oxfordjournals.org/cgi/reprint/163/suppl_11/S1.pdf, Abstract 651, page S163).

ABSTRACT

Epidemiological research is carried out with a final purpose of benefiting public health. Because epidemiological evidence provides information from people about factors harming or benefiting health, the findings of epidemiological research frequently receive emphasis in policy-making. Parties not in agreement with policy directions set with consideration of epidemiological evidence may question the quality of the evidence, often citing inherent limitations of observational evidence and even questioning the utility of epidemiological research in general. Critics may cite the limitations posed by confounding and bias, or dismiss epidemiology as “junk science”. This presentation addresses the strategies of such critics, using case examples from environmental epidemiology and approaches for response. There is utility to the tension generated by critics of epidemiological research, but responding to and countering points of criticism may be burdensome for researchers, particularly those working in areas where epidemiological evidence is given weight in policy matters. Strategies for addressing criticisms range from responding to published letters or other critiques to formal, independent assessments of data quality and re-analysis.

LECTURE (selected portion)

I am going to describe for you some of the framework within which we use epidemiological evidence often in evidence based processes and how across that process there are a range of points where criticisms may be advanced by critics who are stakeholders in the ultimate decision. And what I am going to describe for you comes out of my own experience, watching policy making processes that use epidemiological evidence over the years.

Now here is a scheme that I use to torture students when they are presenting their doctoral thesis proposals. And I ask them the ‘Well, so what?’ question which was posed to me years ago by a colleague in New Mexico. She said, “Well, is this one of those ‘Well, so what’ studies?” And I scratched my head and she said, “You know, one of those studies when you finish it you’ll say, ‘Well, so what?’” And as we just heard, we are doing epidemiologic research because we hope in the end there is one or more stakeholder groups who will use the evidence to advance public health. And if we have done a study doesn’t pass the ‘Well, so what?’ test, somewhere down the line in this scheme we may find that in fact there is no policy utilization apparatus that really cares about what we have found.

So since it’s a long course from having a scientific hypothesis, carrying out a study, publishing the results, and then seeing the results utilized, perhaps cranked into a systematic review as study number seventeen. There your study will be in a table or perhaps you will have a single more significant study that happens to some of us, sometimes. We get more than perhaps a small delta of evidence from years of work. Now how do we use evidence to make decisions? This is a book that actually doesn’t exist, although maybe it should. Now, the process in a very simplistic way that might involve utilization of data to make evidence based decisions is shown here. We carry out studies that generate data from which we hope to learn evidence on associations. We synthesize evidence using one or more processes. One process is for somebody to sit down and read a lot of papers. Another is to assemble an expert panel. Another is to use the systematic review process, perhaps carrying out a quantitative meta-analysis or even doing what I would call pooled analysis, gathering data at the individual level.

We evaluate the evidence that we pull together using perhaps the causal criteria that were proposed by Hill, 1964 Surgeon General’s report, and elsewhere. We use expert judgement and I’m sure many of you have sat around tables drinking coffee and making expert panel decisions. Now the good thing about those exercises is you have to come to a decision or else you don’t get to go home. So sometimes over the course of a few days, stuck away in a hotel room, a hotel somewhere, we make expert panel decisions. And you perhaps come up with things like the evidence sufficient, insufficient, or, every epidemiologist’s favorite, more research is needed, send money please. We use words like weight of evidence to describe how we synthesize. Coming out of that we may take a decision the evidence is sufficient to ensure causality. The evidence is sufficient to infer that trans-fats adversely affects cardiovascular health and levels should be reduced in foods and foods should be labeled. And then from that follows actions.

So this a very general linear scheme that sort of represents what happens in terms of evidence utilization and of course it's iterative and these are arrows going back and forth. But, here is an idea I just want to leave you with because I'm going to show how within this framework there are points where epidemiological evidence, if used, will inevitably be questioned. So here are those points where we might start at the top. Epidemiology is useless, it's juke science. The data are of inadequate quality. Bias and confounding are insurmountable under any circumstances. And even though you've done your best to control for all the cofounders, there is either the possibility of residual confounding or even worse there is some confounder out there that you should have controlled for but you didn't even know it existed. This is an argument I refer to as the "unknown confounder" argument that is hard to beat. What can you say? Maybe there is a confounder out there that has yet to be discovered, but so what. Flawed methods!

So we just heard praise of systematic reviews and perhaps quantitative meta-analysis, but there are those who would say, particularly for observational data, "garbage in, garbage out." And there are a few important people whose points of view on this matter are pretty irrefutable up until the last approximately 20 years while we've been carrying out these full-length exercises. The methods for evaluating evidence may be criticized. The causal criteria are not objectified. They are subject to the opinion of experts. They are not necessarily repeatable. And are experts truly neutral? Or do we all come with our points of views and we reevaluate evidence with a neutrality that some think we don't have? And then on matters such as causation, again, can epidemiologic evidence establish causation or not?

So to criticize an epidemiological study you need to do no more than take an EPI 1 course or equivalent and then sort of turn everything around. So how do you deal in fact with chance? Well to have a powerful study you put evidence together. To have a less powerful study you just stratify until whatever association you are interested in is no longer statistically significant. Misclassification inevitably is there and perhaps its consequences have been to mislead the epidemiologists. Selection bias, perhaps one of those poorly understood forms of bias, can always be conjured up as an unknown or lingering factor. And confounding is the favorite of critics because, of course, there maybe a suite of relevant confounders depending on the outcome. And although criteria for true confounding may not be met, it is always possible to raise the possibility of confounding. And since most people don't really understand it anyway, it's an effective argument in many of these expert panel settings. And then cause criteria can not be met or they are not met. And if you have never had the experience of trying to line up evidence against perhaps Bradford Hill's criteria, it's an interesting exercise to have a panel do so and realize just how challenging it is and how reluctant a group of experts really are to sit, line up the evidence, and then write down sufficient, suggestive, inadequate, and perhaps suggestive of no association. That's not something we like to do.

Now what I want to do is show you one example of these points of criticism of epidemiology and one where over the last 20 years approximately or more there has been substantial criticism of the evidence on passive smoking and disease. And I'm going to use the example of passive smoking and lung cancer in particular. Now here is my scheme I showed you before. The data on passive smoking and lung cancer coming from epidemiological studies, toxicologic studies. And clearly the evidence on active smoking provided some prior basis for

being concerned that exposure to second hand smoke might cause lung cancer. One important body of epidemiologic evidence was on smoking by spouses at risk for lung cancer in never smokers. The technique of meta-analysis has long been applied to this data. Right now the pooled estimate from probably over fifty studies at this point, somewhere around 1.2. The conclusion based on expert evaluation of the evidence has been that passive smoking gives rise to some cause of lung cancer.

And of course the actions that have followed have been to reduce passive smoking with diverse policy and educational interventions. Now just to show you how the points of criticism that I showed you generically have been played out, certainly in this instance we have seen attempts to suggest that epidemiology may not be useful for studying this or other so called weak associations and then confounding and misclassification are substantial concerns. You probably all heard of “junk science,” a term that sometimes arises in relationship to epidemiology. This book probably gave rise to junk science in a formal way most earlier, written by Peter Huber at the Manhattan Institute, a lawyer. Junk science is the mirror image of real science, with much of the same form but none of the same substance.

There are many who dismiss epidemiology as a soft science as opposed to a hard science, which I construe as one which involves using a water faucet. And you can't use a water faucet in epidemiology. In any case, this book is important. And there are entities out there that have the purpose of suggesting that epidemiology is “junk science.” Here is the “junk science” website: www.junkscience.com. Myself and others are regular features on the “junk science” website. And not surprisingly, as the tobacco industry documents have been reviewed, links have been shown to the industry to supporting this website.

In other efforts to suggest that epidemiology may not be adequate, here's the confounding argument in a letter to the New England Journal in 1990 Ragner Rylander brought up the possibility that exposure to environmental tobacco smoke, secondhand smoke, may be confounded, a reasonable suggestion. And in fact an editorial by Marcia Angel appeared in the same issue pointing out that confounding and methodological issues are of concern in looking at so-called weak associations. All reasonable, but what was not revealed was Ragner Rylander's long standing connections to the tobacco industry. So here was a link in which he was attacking the method without showing his own interest in the matter.

In terms of the criticism of epidemiological studies, Hirayama's 1981 publication, one of the first to show that passive smoking increased lung cancer risk, is a textbook. If you read the British Medical Journal over the year following, letters to the editor were written that essentially covered the gamut of EPI 1 criticisms of an epidemiological study. It's a useful teaching exercise to go through those letters, where the issues of confounding and misclassification were raised. There were some rather difficult and subtle issues in misclassification that merit serious attention to those studying this association. Malcolm McClure gave us a nice framework for thinking about bias and when we look at this association, secondhand smoke and lung cancer, confounding, information bias, the possibility that only positive studies were published, have all been brought up, and again, not unreasonable. When Hirayama's study was published there was a question as to whether there was a statistical error. Nathan Mantel was brought in as a consultant to the tobacco industry and indeed a problem was

identified in the calculation of the test, the statistic, but with correction, the finding remained statistically significant after research was carried out. This is a complicated story that I will always say that here's the end result publication to address some of the subtitles of classification in this particular paper. So, here is one paper with a number of interested stakeholders (the general public, public health, the tobacco industry, and others of course) that went through the full gamut of criticism.

Other concern in this story is the question whether we know about mechanisms and of course biological probability is a critical component of assessment of potential causality associations and here issues are raised about unknown mechanisms and whether there was, in fact, an animal model. Here, for example, is the mechanism argument. This is me being questioned in the Minnesota tobacco trial. Isn't it true that notwithstanding the amount that science may know about the P53 gene, that science cannot yet set out step by step by step (at least three steps I guess) the mechanism by which cigarette smoke causes lung cancer. And was interesting today hear Marcia Angel make a comment, in fact, to this point that we still don't know the mechanism.

Which raises the question, when do we know the mechanism? And what are we talking about? And can that be a criticism inevitably raised in trying to pull out the biological plausibility support for associations observed in observational data? I suspect that decades from now, we will, in fact, know substantially more about the mechanism by which tobacco smoke causes lung cancer. We will be defining it at new levels of molecular precision. Do we know the mechanism or have support for plausibility? And of course comes the question, why do we need to know the mechanism? For those of you who can't see the bottom, it says these studies are inconclusive. So far we have only succeeded in giving cancer and heart disease to laboratory humans. And with 1.1 billion smokers in our human laboratory now, do we need animals? The fine print over there says what do you do when you run out of humans: start using lawyers.

Continuing down the scheme, meta-analysis has been criticized. The Environmental Protection Agency used meta-analysis in its 1992 risk assessment and again here criticism surfaced on meta-analysis, quantitative meta-analysis, as a method for summarizing evidence. And again, you know in terms of sort of spotting critics, when you see papers with titles like these supported by a major stakeholder you realize, of course, it may come with a possible point of view. Another way to attempt to set aside evidence is to propose alternative criteria for causality, to put together panels to also review the evidence. This has also been done.

There is a long history in use of epidemiological evidence as to propose, in fact, that epidemiological evidence, in combination with the other lines of evidence that we consider, are not adequate for proving causation. This is one of the arguments advanced in the case of both active and passive smoking for a long time. This quote comes from the lead attorney for the tobacco industry in the Engle case in Florida: "Science requires three things in order to get to causation: First, statistics (I have that from the epidemiological studies); second, animal studies (which follow-up on hypotheses suggested by epidemiological studies); third, understanding the mechanism (you can see this elegant definition: 'the mechanism is the thing that causes it')". And apparently we don't quite yet have this thing.

Now, what I have tried to illustrate for you is some of the tactics that are used. And now I'm going to suggest that you need to begin to think if your work, or work that your looking at some evidence synthesis process that may lead to policy, may invite critics. So, I think epidemiologists doing studies that are relevant, which hopefully is most of the work, undoubtedly will be producing evidence that will be of interest to one or more stakeholder groups. So carrying out studies that are important, which I hope we all are, then there will be people who are interested. And they come from multiple points of view and if your evidence is likely to be used in a policy making process it is likely to be of great interest. You know, perhaps you are going to be study number seventeen on a forest plot. But its better than not being in the forest plot at all and you're making your contribution.

There is also what has happened in the past and I know if I publish in certain areas that undoubtedly my work will be looked at by certain people because they have been doing it for years and I think again for those working in a particular field, you may come to know your sort of critics, if you will, who are working for stakeholder groups that are going to approach you and either tell you how much they like your work or possibly how stupid you are. And this is also inevitable. And just as one example in our work on air pollution there are certain people we often hear from. One for example is Suresh Moolgavkar, a useful critic, but someone who for the last ten years our air pollution group has interacted with. And here is one of his comments in a recent paper related to a very important public health issue, the fine particle standard.

So there is a range of people who will criticize and I think that the people who I see in different policy making settings range from very thoughtful critics and skeptics, who keep us all honest, to people who may come with a point of view and if you may want to call them hired guns, so be it. Some may have a track record of being spokespeople for the stakeholders on a particular group and some may not. And inevitably the tactics follow some of the lines that I have showed you in terms of addressing whether studies are or are not informative.

Now what do we do to anticipate criticism? Well, we do good studies. And I think for those who are doing studies that are likely to be important in the policy realm, you need to be prepared to say that you have done a good study and to prove that you have done a good study. Increasingly that means that you have well documented agency protocols; you may have had an external review, if it's a particularly important study and you can produce the protocol; you've done a transparent analysis; and you have carried out relevant sensitivity and certainty analyses. And this may be very critical in assuring the results are viewed as important and unbiased. But for the problems of weak effects and some of the other methodological problems, I also think we have things in the epidemiological tool box. We need to do studies that are large enough, possibly a pooling of evidence, either at the individual level or carrying out quantitative meta-analyses.

We are in this new world where there is an expectation that we will share our data. In fact on certain matters, for those of you familiar with the Shelby Amendment, you will be aware that there may be a legal requirement that we share data if the evidence that we have comes from federally funded research and it has figured into a government decision, regulatory decision with policy implications.