

## **Our estimates *are* uncertain, but that's ok**

Carl V. Phillips, MPP, PhD  
University of Texas School of Public Health  
Houston, Texas  
cphillips@sph.uth.tmc.edu

June 2000 (reedited February 2002)

**ABSTRACT:** Epidemiologists and other health researchers act as if only precisely stated point estimates can be used in decision making, despite policy analysts having tools to make good use of estimated uncertainty. Epidemiology methods are thus structured to provide only these estimates and to overestimate precision. But health policy decisions should be based on expected net value, which can only be determined when we admit to and quantify our uncertainty. The standard practice, to stick with the status quo until an alternative is "proven" to be better, falls far short of this gold standard, forcing various ad hoc patches. It also makes epidemiology relatively less influential compared to other methods, such as toxicology and clinical trials, which are much more precise but not necessarily accurate measures of real-world exposures. We need to recognize that it is possible to quantify the inevitable uncertainty in epidemiologic research and that policy should be made with explicit recognition and measurement of that uncertainty, rather than reporting only test statistics that pretend we are more certain than we actually are. Only after recognizing this, we will have the incentive and opportunity to improve our technologies for estimating uncertainty and using it, allowing epidemiology to provide its best possible contribution to health decisions.

**Keywords:** epidemiology methods, health policy, uncertainty, cost-benefit analysis (CBA).

To make optimal choices about health, decision-makers need point estimates of their parameters of interest, plus an honest and effective description of the total uncertainty surrounding those estimates. Health researchers do not provide this, but instead typically report point estimates and simple summary measures of random uncertainty (p-values and confidence intervals). This fails to quantify uncertainty caused by nonrandom errors, and fails to report the full amount of information about uncertainty that decision-makers can utilize. Policy makers need information about the full distribution of possible values of the parameters of interest due to all sources of error, both random and systematic. This information can be extracted from studies by adopting available methods, so addressing this problem is a matter of changing what we choose to report.

A second problem runs deeper. Attitudes and methods in epidemiology and other quantitative health research are biased -- due to accidents of history and other forces -- toward creating the limited information that is (mistakenly) perceived to be most useful for policy making, even though such reporting is not an honest representation of what we know. Addressing this error will require attention to the optimal methods for making tough policy decisions and an active effort to feed that process the information it really needs. Until this occurs, epidemiology will fall short of its true potential contribution to scientific knowledge while simultaneously overstating its contribution by overselling the certainty of results. That is, epidemiology methods fail to analyze the underlying uncertainty that they could measure and that policy makers could use. Instead, epidemiology reports results that incorrectly imply that the uncertainty is low and the point estimates are extremely precise.

The failure to quantify and report total uncertainty results in a vicious circle. Researchers misperceive the needs of policy makers. We fail to quantify total uncertainty, thereby underestimating

confidence intervals by reporting only random sampling error. We seem to mistakenly think that low p-values and narrow confidence intervals are the most important decision parameters. Meanwhile, policy-makers misperceive what researchers can provide. They expect only point estimates and do not realize that they can ask for estimates of total uncertainty, both systematic and random, while mistakenly believing that the reported point estimates are sufficient without such quantification because they are portrayed as much more precise than they are.

The following looks at the economics- or policy-analysis-based decision making process and the information needed, and explains that researchers can provide this information, simultaneously improving policy decisions and bolstering the importance of epidemiology.

#### Health Policy Decision Making, Optimized

Health policy decisions (defined broadly, to include widespread behavioral recommendations and setting professional standards of practice, as well as government regulations and other formal public policy) will always involve some costs and some benefits that must be traded off and some uncertainty that should be formally considered. For making good health policy, it is not sufficient to merely know whether Treatment A is more often effective than Treatment B, or whether Exposure X increases the chance of Morbidity Y. We need to know how much better or how big an increase, or, more precisely, *the probability distribution of our best assessment* of how much better or how big. In addition, we need to know the relative costs of A and B, of suffering Y, and of make the effort to avoid X. Unfortunately, many of these key parameters are never addressed.

It is widely accepted in most theories of public policy, decision theory, and applied economics

that the expected net value of an action should play an important role in assessing whether it is good or bad.<sup>1,2</sup> Most decisions will ultimately be made in a social and political world rather than following directly from the specific conclusion of quantitative analysis, of course. But this does not excuse analysts from providing the best possible scientific advice, even though we know that policy makers will be influenced by other factors also. Moreover, the role of the scientific advice in shaping these other influences should not be overlooked; advocates and politicians will take many of their cues from the research, or at least will base many of their arguments on it.

Despite the need and importance of decision-theory-type analysis, badly-formed rules of thumb and standards of practice dominate the discussion of quantitative matters in public health policy, and the basic lessons of economics or decision theory are lost. The dominant rule of thumb in practice can be summarized as "don't do anything (i.e., keep doing whatever we are doing) until we are sure that something else is better, whereupon switch to the alternative." This rule is part of the vicious circle. Policy makers expect "yes or no" answers from health researchers, and design policy accordingly. Researchers recognize that only point estimates will be used and only statements of certainty will affect policy, and so report accordingly. Worse, we have structured our methods to provide only those statistics.

To escape this trap, we need to start by realizing that,

- an optimized decision can be made based on whatever information is available,
- and, indeed, a decision always *will be* made no matter the quality of the current information.

To keep doing what we were doing yesterday *is* a decision like any other. But the bias favoring

it over other decisions is strong, as is demonstrated in epidemiology textbooks, medical research handbooks, policy guidelines, and the actual decisions made. Rather than assessing the expected net benefits (or, equivalently, loss function) of different options, it is typical to continue current policies until an alternative course of action passes some statistical test one or more times. This makes an implicit (and frequently incorrect) assessment that the currently available information tells us that the status quo is the best decision.

A valuable tool in decision making is the net expected value of a state of the world, determined by integrating (or summing) the net value of a policy across possible values of the parameter of interest based on our current knowledge (however imperfect) of parameters of interest. That is, we want to determine the expected net value,

$$\int f(x) (\mathit{Benefit}(P,x) - \mathit{Cost}(P,x)) dx \quad (1)$$

and compare it across all policy options. The *Benefit* and *Cost* functions are the social impacts for policy option  $P$  given state of the world  $x$ , the parameters of interest. Function  $f$  is our probability density on the parameter of interest,  $x$ , given current knowledge, which (like the integral itself) is a notational shorthand, since  $x$  will generally be multi-dimensional and complicated. The "net" refers to the subtracting of costs and "expected" to the probability-weighted average of outcomes.

For example, the parameter of interest might be the relative risk of cancer due to exposure to a pesticide, with one policy option being banning the pesticide, which has the possible benefit of reducing cancer by a given amount (as captured in  $f$ ) and a cost of the value of using the pesticide. If decisions are based on utilitarian assessments of costs and benefits, we should choose the policy with the highest

calculated value based on available information, even if it is not "proven" to be better than the status quo. But even given that public health decisions are seldom based on purely utilitarian calculations, the calculation is a useful input, often demonstrating such a huge imbalance that it is difficult to argue it should not be decisive.

The net expected value calculation may result in a policy that has a substantial (even majority) chance of being inferior (as measured by the sum of benefits minus costs) if its net benefits are high enough for some portion of the density of  $x$ . This is a major departure from our current emphasis on certainty. Real-world loss functions from making suboptimal policy choices are seldom symmetrical about the estimated mean of the parameter of interest, and so knowing whether that mean produces positive or negative net value is not sufficient. A policy that is probably a bit better than the status quo, but has a small positive chance of being a disaster would be shown to be inferior.

For example, a chemical that provides a bit of benefit compared to the alternatives but might be causing widespread reproductive damage would have a net negative expected value, even though it is impossible to "prove" it is unsafe. A new policy could displace the status quo even if there is a substantial chance that it actually has lower net benefits, as long as the net losses are low should they occur, while the net benefits are high. Compared to this gold standard of expected net benefits, the status-quo-biased process can offer the wrong recommendation either in the direction of doing too little to protect health or doing too much (because the benefits do not justify the costs).

To partially make up for the departure from using expected values across our uncertainty, we have introduced kluges, devices intended to patch around the problems with the underlying methods rather than to fix them. These include various versions of the "precautionary principle" that are often

invoked (though seldom defined) in advocacy and policy making. These principles, which generally preclude policy decisions that create health or environmental risks of unknown magnitude, offer a rough substitute for correctly summing in a tiny chance of disaster. Other makeshift precautions include rules that exposures be ten or a thousand times lower than the level of exposure at which negative impacts have been "proven." This substitutes for an explicit consideration of the expected net value implications of uncertainty. Like the status quo bias itself, these kluges sell short the potential contribution of epidemiology: we do careful calculations and then inflate the results by arbitrary factors.

There is no question that doing the calculation instead of using the rules of thumb is difficult. Indeed, calculating the probability of an unexpected disaster strains the limits of our imagination (by definition), let alone our ability to quantify. This does not, however, change the gold standard and the benefit of pursuing it in our research design and analysis. The impact of health research is frequently many lives and many dollars, and the research itself is expensive, so it is impossible to justify not even *attempting* to estimate the missing values. When we fail to quantify, decisions are wholly given over to a political process, and we are doing science and policy no favors by being too modest about our ability to quantify.

### Epidemiologic Data for Decisions

As daunting as it seems, epidemiology can estimate  $f(x)$ .

Epidemiology offers the possibility of measuring exactly what we are interested in -- such as the effect of a real-life exposure on a variety of people in an everyday setting -- with all the variability and uncertainties introduced by real people in the real world, and the possibility of measuring the variability

and uncertainty themselves. Other health research technologies usually offer more precise measurements of substitutes (possibly not very good ones) for what we actually want to know. Toxicology offers us very controlled and precise measures. But since mice receiving a huge dose of toxin in a controlled setting are seldom our target population of interest, it is obvious that epidemiology offers some major advantages. Less facetiously, the correlation between results of animal experiments and the effects on humans are fairly poor (thus our large safety margins). Precision should not be confused with accuracy. Randomized controlled trials theoretically allow us to measure anything with great precision, but great expense and ethical concerns dramatically limit their use. (Furthermore, RCT research could also benefit from a better understanding of how to quantify multiple sources of uncertainty.)

Population studies are the obvious approach to measuring what we really want to know. But despite this huge advantage of observational epidemiology, it appears that the field suffers from "toxicology envy" (much like old-school economists are often accused of having physics envy), and so surrenders its advantages while trying to willfully ignore its weaknesses. Environmental epidemiologists are sometimes loath to admit the true uncertainty in their results because they believe it will give even more policy influence to toxicology with its great precision. Indeed, as long as we allow health policy discussions to focus on being able to reject *some* null hypothesis that is *somehow* related to the policy in question (even rather tangentially), observational epidemiology -- despite its unquestionable contributions -- will remain a poor cousin to bench science and the occasional RCT.

Epidemiology, with its many sources of uncertainty, is simply not good at providing evidence that something is true beyond a reasonable doubt. With notable exceptions like the link between

smoking and lung cancer, population studies seldom demonstrate causal effects that we can believe beyond any reasonable doubt. Instead, they produce a range of plausible values. Since decisions are necessarily made, and should be made based on our assessment of what seems to be better -- rather than what is Proven To Be True, this is ok. Eliminating uncertainty would be the ideal solution, but given that it is not feasible, quantifying the true uncertainty is a useful alternative. Understanding the true uncertainty does not detract from the value of epidemiology, it enhances it.

The tendency to devote our research to finding associations where we can reject the null hypothesis (using our incorrectly narrowed confidence intervals) contributes to data-mining (to create "provable" hypotheses based on the data) and publication bias (reporting only the "proven" results). The resulting barrage of contradictory claims damages the credibility of health research. Researching the effects on overall mortality and quality of life resulting from common decisions (common environmental exposures, major dietary components) is useful even if the effects are uncertain. But the emphasis on rejecting some null hypothesis creates the incentive to dredge for possibly random links between very specific exposures and particular diseases where the uncertainty can be obfuscated by the p-values. Clearly science and the public health benefit from learning about such associations when they are real. But we apparently forget the lesson from first-semester statistics, that trolling for  $p=.05$  will generate an article from almost one out of twenty pairs of random number series (or more, if multiple functional forms are tried) if we do not account for multiple-hypothesis testing. As we search for statistical significance as a goal in itself, and overestimate confidence, results pour into the scientific and popular press. The barrage of conflicting narrow results, reported with certainty, exhausts non-scientists' willingness to think about health-affecting decisions and their trust in health research.

## Getting the Right Density Function

The next step after recognizing the importance of properly reporting uncertainty is to properly measure it. The technology for this in its infancy,<sup>5-7</sup> but the basic idea is simple. Most research currently practiced implicitly pretends that only random error creates uncertainty in the measured results. Of course, no one ever states this obviously inaccurate claim, but the reported summary statistics (confidence intervals, p-values, etc.) are based on this implicit assumption. Other sources of error are mentioned qualitatively, and then generally ignored. As a result, the other sources of error are completely lost on most policy makers and the public. They read the point estimate and that it is "statistically significant," and take the apparent causal relationship as fact, even when experts who read the literature know that the various sources of error mean that in most cases the results are quite speculative.

Any research method, including clinical trials and bench research, involves some uncertainty beyond random error, and that uncertainty should be quantified. Population research (including epidemiology, econometrics, and other fields) faces more and generally larger sources of error. There are errors in the measurement of both exposure and disease, uncontrolled confounding, non-response, losses to follow-up, missing data, and a myriad of challenges with model specification. Many of these can be quantified, and we can certainly do better than just assuming they have no effect or an attenuating effect. Only by recognizing that this is useful and possible will we take the steps necessary to get good at it.

## Conclusion -- A Call for Action

We need to recognize that there are methods for quantifying multiple sources of uncertainty, that epidemiology inevitably involves such uncertainty and can measure it in useful ways, and that policy should be made with explicit recognition and measurement of the uncertainty, rather than the fiction that we are certain. Only after recognizing this, we will have the incentive and opportunity to improve our technologies for carrying out each of these steps.

Epidemiology must learn to embrace and properly respond to the inevitable uncertainty, lest it surrender too much of the study of human health to those who can precisely measure things we are not actually interested in. To become a more honest science, epidemiology needs to start accurately reporting the substantial uncertainty in its results. To simultaneously keep itself positioned as a critical science in health policy making, it needs to recognize how to properly use known uncertainty in decision making and to educate others on that point.

## REFERENCES

1. Stokey, E and Zeckhauser RJ (1978). *A Primer for Policy Analysis*. New York: Norton.
2. Friedman LS (1984). *Microeconomic Policy Analysis*. New York: McGraw-Hill.
5. Phillips, CV and Maldonado G (1999). Using Monte Carlo Methods to Quantify the Multiple Sources of Error in Studies. (Abstract.) *American Journal of Epidemiology*, 149(11): S17.
6. Phillips, CV and G Maldonado (2000). Generating Fully-Articulated Uncertainty Distributions for Epidemiologic Studies. Unpublished.
7. Phillips CV (in press), The Economics of "More Research is Needed". Forthcoming in *International Journal of Epidemiology*.

## ACKNOWLEDGMENTS

The author thanks George Maldonado and Karen Goodman for helpful suggestions.