

The Economics of "More Research is Needed"

Carl V. Phillips, MPP, PhD

University of Minnesota School of Public Health
Division of Environmental and Occupational Health

phil1047@tc.umn.edu

See also Posters #161 and #162. Suggested order for reading: 162, 161, 160.

Manuscripts available for 161, 162, and half of 160. The other half of 160 (food safety) will be available within a few months.

The author thanks George Maldonado and participants in several University of Minnesota School of Public Health seminars for helpful suggestions.

"More Research Is Needed"

Studies in epidemiology (as well as many other fields) conclude with the mantra "more research is needed" so frequently that the phrase is commonly used as a humorous catch-phrase.

How are we to interpret this statement?

It says more than the neutral hedge, "we do not know enough to draw definitive conclusions."

"We do not know everything yet, so could learn something more," is vacuous.

The best interpretation is the economic statement, "the expected benefit that would come from more research, due to the resulting improvement in our estimates of parameters of interest, justifies the cost of that further research."

This suggests that some assessment should be made about the tradeoff, but the economic statement is virtually never accompanied by economic analysis. It is difficult to understand how applied research can be justified at all, let alone be declared to be "needed," without such analysis.

Why Worry about the Costs and Benefits of Epi Research?

Epidemiology is not an abstract academic exercise.

Our choices of research affect:

- C medical treatment options,
- C lifestyle choices,
- C millions of life years,

- C expensive regulations and production methods,
- C billions of dollars of the economy,

- C the use of millions of research dollars,
- C the availability of the limited pool of skilled researchers for other important projects.

Other areas of endeavor (most with much less important consequences) continuously invest in assessing current and future activities, spending at least a few percent -- often a lot more -- of the total budget on figuring out if the rest of the budget is being spent well.

This is often done for health interventions¹ and pharmaceutical research, but is badly overlooked in other areas or health research.

Before embarking on an expensive study whose results are intended to inform policy (including formal public policy as well as public health officials' recommendations about lifestyle choices, standards of practice in medicine, and the like), it is natural to ask *how* the outcome is likely to inform policy.

There has been little or no formal analysis of when a particular epidemiologic study is likely to be worthwhile, let alone applications of such analysis in study design.

It is not easy,
but it can and *should* be done.

If we do not take full advantage of past research when designing future research, then why did we do the past research?

Contrary to the impression the public gets from health news headlines, the research process is one of building upon existing knowledge rather than stumbling around until the definitive result is found and displaces all previous findings.

Yet further research is often done as if researchers believe the headline version of scientific progress. Typically, research simply repeats existing studies, possibly correcting existing errors and possibly not, as if the next study will be definitive.

Calculating the Value of Further Research

Basic economic principles:

Nothing is free, so the possibility of some benefit is not sufficient to recommend something.

Decisions should be based on an assessment of expected net value.

That applies to any decision, including decisions to do more research on a subject (and the choice of research projects among many options).

Net value = costs - benefits.

Expected = probability weighted average (primarily of the benefits).

So, we need to figure out the costs and the distribution of the benefits.

Determining the Expected Benefit of Research

If we can describe:

- C how a subsequent study might change our current estimates of the true value of the parameter(s) of interest,
- C estimate what the probabilities of those changes are, assess how we would change our behavior under the new information,
- C and put a dollar figure on the behavioral changes,
- C then we can compare the latter figure to the cost of gathering further information and figure out if it is worthwhile.

It is possible to create the tools needed to do such analysis.

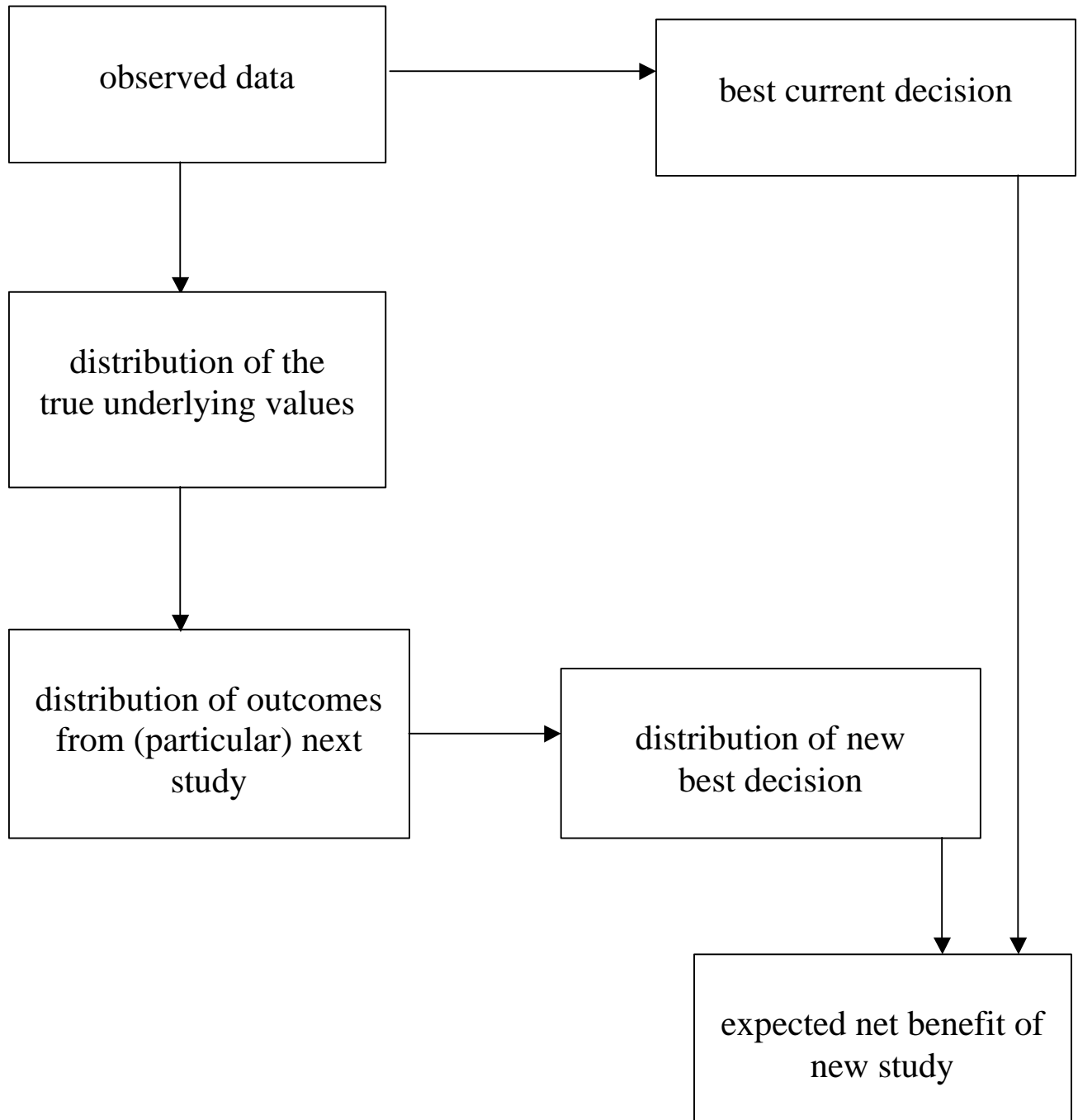
Working backward from the information we would like to have:

1. Ultimately, we want to know what policy decision new information would generate. For purposes of applied research, new information pays off if and only if it changes our assessment of the optimal policy. (The question of science for its own sake is a different matter, and is beyond the present scope. The justification for funding further applied research in epidemiology and similar fields is almost always presented in terms of specific practical outcomes, and so should be judged on that basis.) The value of such a change can be calculated based on the net improvement in outcomes under the new policy, given our (updated) assessment of the parameters of interest.
2. Naturally, we do not know the answer from 1. ex ante. But we can identify the possible results of the new research and how those results would change our assessment of the parameters of interest, and thus what new policy decisions would result.
3. To quantify the implications of 2., we must assess the probability distribution of the outcomes based on our existing knowledge. This is a quantitative and philosophical challenge, but it is clearly possible to generate better information than we do now.

Reversing the steps, the probability distribution of the true value of the parameter of interest is calculated based on current knowledge and broken down into portions of the density that correspond to various findings that could result from the next study, each of which has a particular impact on our understanding of the world and thus policy decisions. The probability distribution of those results can be generated, and the expected value of the information compared to its cost. (See Figure 1)

Figure 1

Process for Calculating Value of a New Study



Formal Statement of the Optimization Calculation

Consider a case where a policy decision, POP , is made based on all presently available information, denoted by XOX . We want to choose P to maximize the expected net benefits from a policy, $B(P,X)$. Define,

$$R(X, X_T) / B(\underset{P}{\operatorname{argmax}}(B(P,X)), X_T), \quad (1)$$

the realized net benefit from choosing the optimal policy based on the belief that X is the state of the world when really X_T is the true state of the world. Then if further research, study s , allows us to update our belief about the world to X_s , the true net benefit of that research is,

$$NB_s = R(X_s, X_T) - R(X, X_T) - c_s, \quad (2)$$

where c_s is the cost of carrying out the study. We are never going to know the true value, X_T . Nor will we know X_s before doing s . But we can always make the best possible prediction of their distribution based on existing knowledge.

To estimate that expected value of doing s , we calculate expected benefit based on our current beliefs about the distributions of X_T and X_s given X . For the simplest case, assume that the states of the world are continuous scalar values (such as a single relative risk). (Other cases follow by analogy, requiring we take the appropriate n-dimensional integrals and/or sums of probabilities.)

Then we want to know,

$$E(NB_s) = \int \int (R(X_s, X_T) - R(X, X_T)) f(X_T|X) g(X_s|X_T) dX_s dX_T - c_s \quad (3)$$

where $f(X_T|X)$ is the density function for the true value given our existing knowledge, and $g(X_s|X_T)$ is the density of the results of s given the true value.

With $E(NB_s)$, we can compare the expected payoff of doing a study to not doing it, and compare the expected payoff of alternative studies.

Notice from Equation (3) and Figure 1 that the prediction about the outcome of the next study, X_s , follows directly from our distribution of the true value, X_T , which is based, in turn, on our current data. This line of reasoning resembles Bayesian updating, though the economic analysis is agnostic with respect to statistical methods. There is no requirement to use Bayesian or any other particular method. (The formulation presented here uses no prior probabilities, basing predictions about true values on observed data and predictions about future observations on the predicted true values.)

Determining f is difficult and even g is nontrivial. Indeed, determining these densities is widely regarded as impossible, and thus seldom even considered. However, just as some policy can (and will) always be made given the available data, some best estimate of f and g can (and should) always be made for X . For purposes of demonstrating the value of cost-benefit analysis of further research (and of creating new tools to calculate f and g), it is sufficient to recognize that we are not completely ignorant of f and g .

Stylized Example

- C Exposure to an environmental chemical
- C May increase the probability of a disease.
- C Binary exposure and disease classification.

- C Choice of a single regulatory intervention, changing production processes to eliminate emission of the chemical, costing \$200 million in present value terms and eliminating the risk.

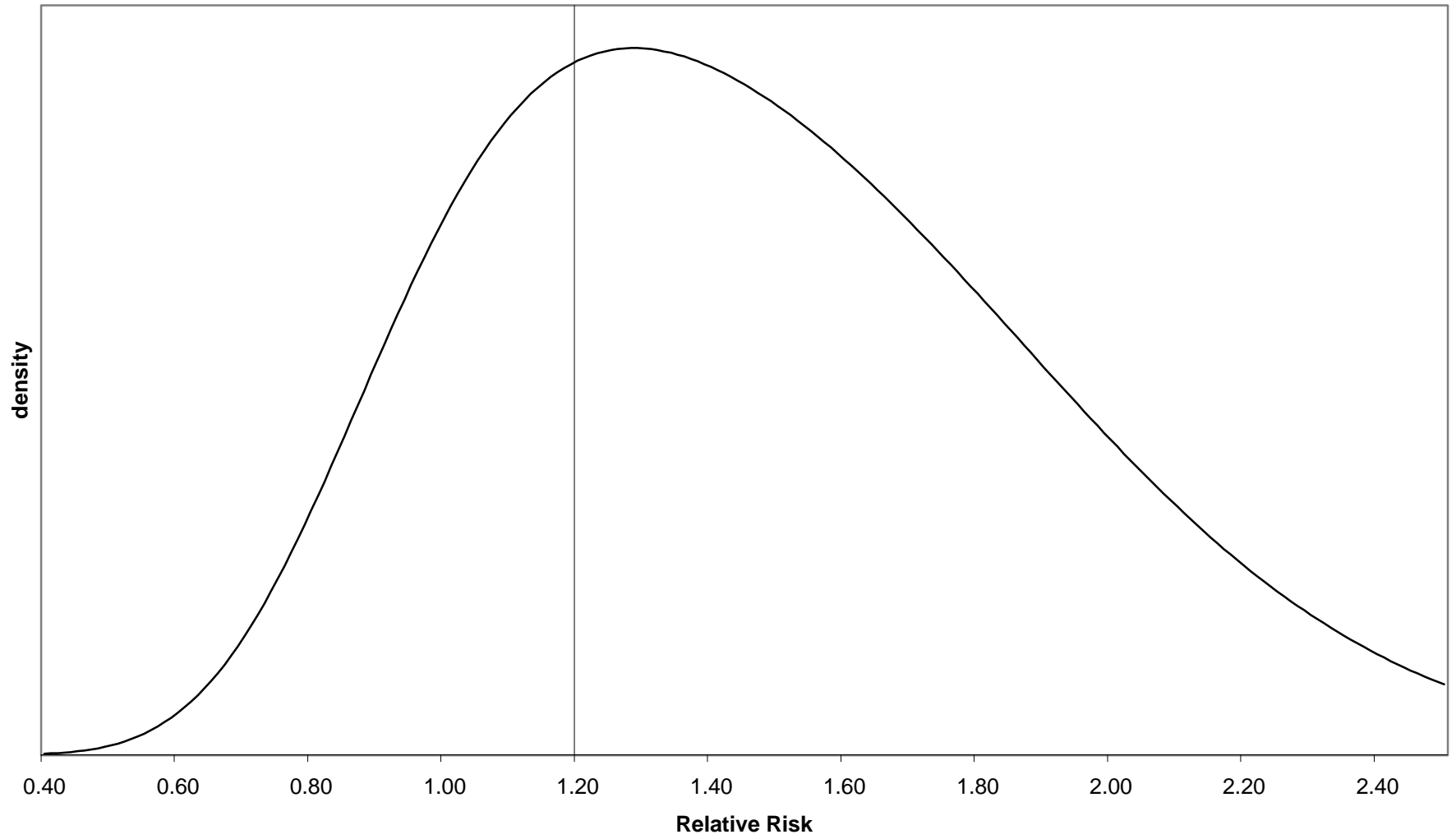
- C Each 1% increase in relative risk (RR) of disease for the exposed population costs society \$10 million in the form of increased morbidity and mortality.

Then, intervention is warranted if the RR is greater than or equal to 1.20, the *action point*.

Assume our existing belief about the true relative risk from the exposure is distributed according to Figure 2, based on the data from an existing study. Intervention is warranted, as determined by taking the probability-weighted average for net costs of the exposure without regulation and finding it is higher than \$200 million.

Notice that we are not concerned with p-values or confidence intervals. As with other true optimization calculations, our policy decision is based on our best assessment of the current data, regardless of how confident we are that this is the right decision. Despite the fact that almost a third of the probability mass is on the "wrong" side of the action point of 1.2, there is still a best possible choice -- to intervene -- and it should be made.

Figure 2



Further Study

New study, s : repeats the existing study for a larger population (specifically, increasing the total sample from 1000 to 3000).

It turns out that the process that generated the distribution in Figure 2 is a combination of possible bias from disease misclassification (in particular Type I error, as discussed below) and random sampling error. (We ignore the other inevitable sources of error to simplify the example, but they would be included in an actual analysis, as presented in poster 161.) Assume that s will not affect the uncertainty about the level of exposure misclassification because it is basically the same study.

Three possible results from s :

1. We could confirm our existing belief that we should intervene, thus not changing our behavior and generating a benefit of zero minus the cost of s .
2. We could discover, correctly, that intervention is not a good idea, with the benefit of the resulting change in behavior depending on the true value of the RR.
3. Or we could "learn" that intervention is not a good idea when it really is, with the net cost of the resulting unfortunate change depending on the true value of the RR.

A Monte Carlo (see poster 161) simulation produces the distribution of the value of the resulting change in policy (without the cost of s) in Figure 3, which shows a probability atom at zero and a density for other values.

Figure 3

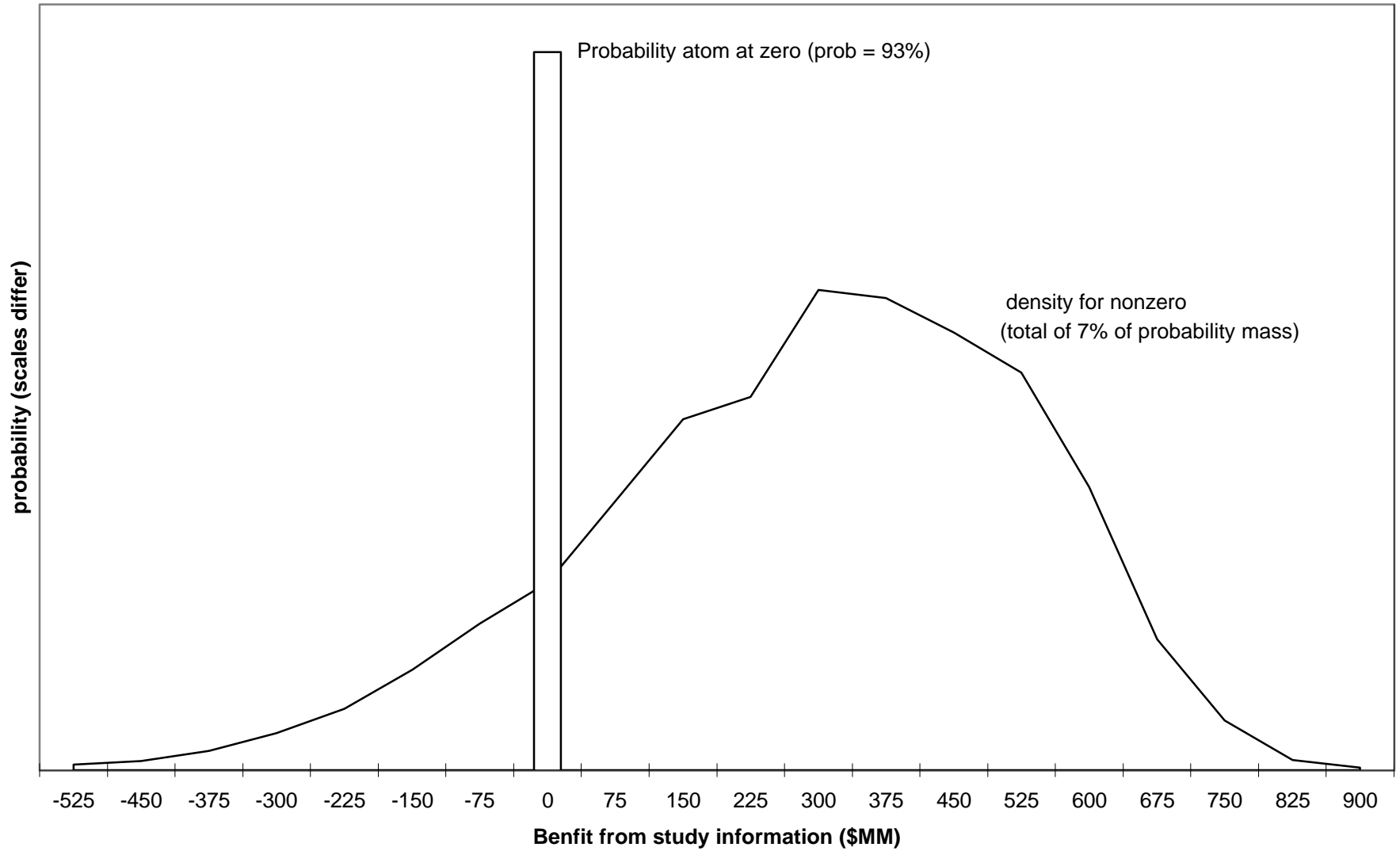


Figure 3 shows that the welfare change has a strong mode at zero, representing the probability of not changing our behavior. The new study has non-zero payoff only if it tells us that our old decision was not optimal. If the study merely reinforces our previous decision, we still get the same actual benefit.

This illustrates a flaw in the standard vision of when to do more research. Under the typical status-quo-biased, p-value-centric model of decision making, the first study, represented in Figure 2, would not have been enough to warrant intervention and a further study would be needed to provide the excuse to take the policy action we already thought was right.

Under the standard paradigm, researchers are searching for a study that will tell them again what they already think they know. This makes poor use of available information by failing to act on information in advance of some arbitrarily-set level of confirmation. Additionally it departs from the hypothetico-deductive scientific method that underlies most modern scientific thinking, in that it sets out to confirm existing beliefs rather than subjecting them to severe tests to see if they hold up.

So, is s warranted?

Applying Equation (3) and numerically integrating the values from Figure 3 yields $E(NB_s) = \$18$ million - c_s . Doing s would likely be worthwhile, though not if it were extremely expensive to carry out. When there is \$200 million worth of productivity at stake, it is worth a lot to make sure there really is a big health hazard.

This sounds similar to the justification for demanding a high level of statistical certainty (a low p-value) before acting. But statistical certainty rules are a poor substitute for the cost-benefit approach. In particular, the standard tests completely ignore the key values in the optimization calculation: the costs of the intervention and disease. If the cost of the disease is low enough, the cost of intervention is either very low or very high, or the cost of the new study is high enough, then further research will not be worthwhile, whatever the p-value. If the intervention were cheap enough we should be less concerned about unnecessary intervention, and regardless of the unimpressive p-value it would be best (from the economic and scientific perspective) to just change our industrial practices based on our strong suspicion and move on.

Consider an alternative further study, v .

The distribution represented in Figure 2 is the average of two distributions illustrated in Figure 4. Distribution X_R represents the distribution from sampling around the actual observed data, uncorrected for misclassification. Distribution X_L represents the researchers' concern that there was a disease misclassification, wherein exposed individuals were inaccurately judged to have the disease, creating a bias factor of about 1.5. There is still some apparent effect of exposure, but X_L would not justify intervention.

Validation study v will resolve the question of misclassification bias, and give us a revised distribution X_v that is either X_L or X_R .

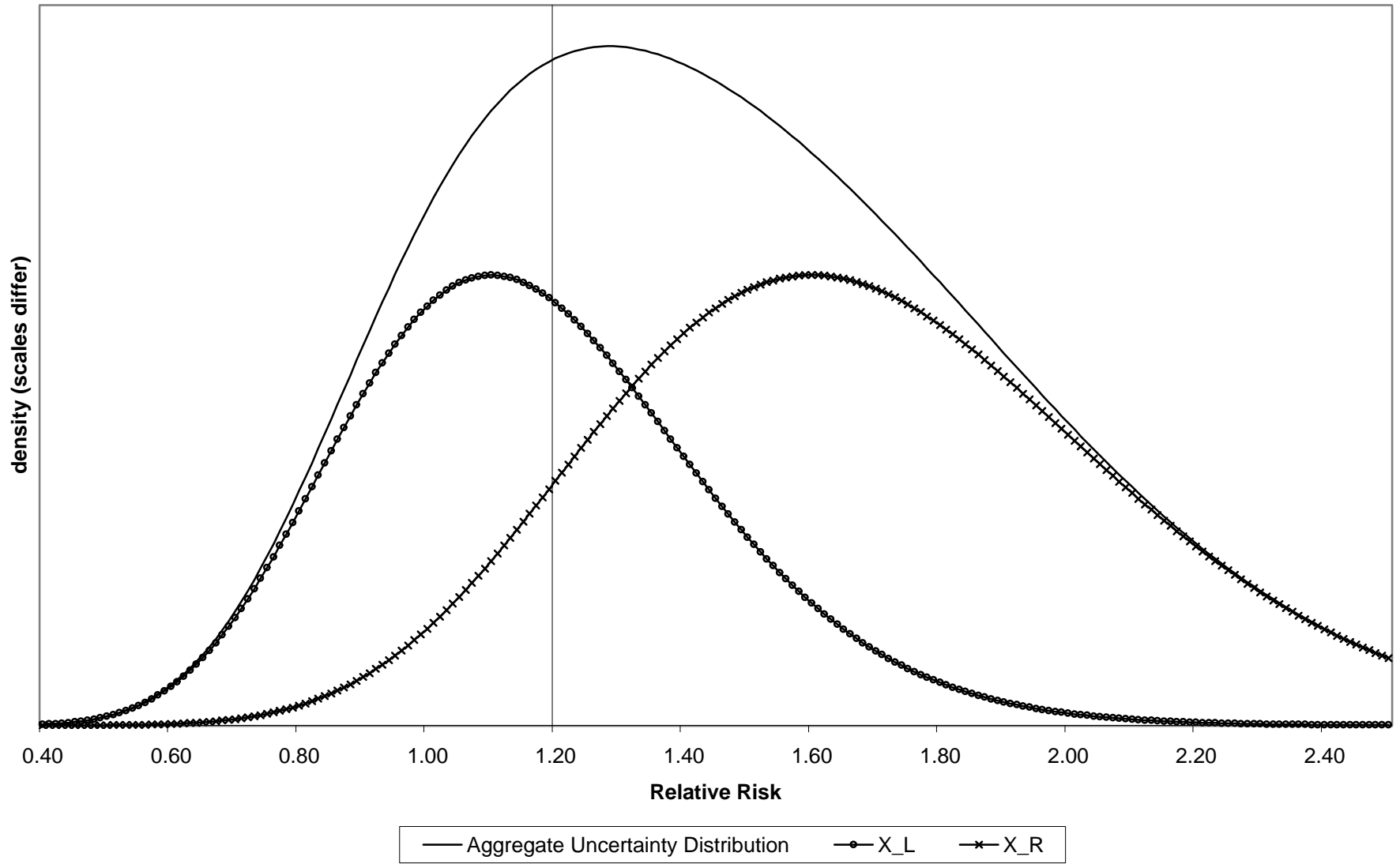
Is v warranted?

The study pays off if we decide to act differently under X_v than we were doing under X . If $X_v=X_L$ we should not intervene, while if $X_v=X_H$, we should continue our intervention.

Once again this produces a result that is contrary to the usual attitude that puts greater value on finding X_H , disclosing a more certain and serious problem. Although it is reassuring to confirm the wisdom of the apparent best action, it lacks the practical payoff of finding out we were wrong.

The value of the study lies in the .5 probability of finding out that the \$200 million expenditure generates an expected benefit of only \$166 million (determined by numerically integrating X_L), an expected benefit of \$17 million.

Figure 4



Practical Implications

It may only take a few person-days (and almost certainly less than a few person-months) to assess whether an expensive, multi-year study is really worthwhile.

Just as researchers are fond of telling policy makers that they should spend at least a few percent of the cost of interventions to assess whether a policy is effective, it is worth a few percent of researchers' efforts to see if their research is likely to be effective.

Epidemiology curriculum does not recommend it or even suggest it is possible. Journal articles do not report it having been done, even when high stakes policy relevance -- not to mention the explicit calls for further research -- cries out for it.

Simply calling attention to this way of thinking could be quite beneficial, even if few formal calculations are done. It may be infrequent that a validation study has a .5 chance of saving \$200 million, but when the possibility presents itself, everyone should have the basic tools to notice.

Are these analyses actually possible?

How can we possibly calculate the densities, $f(X_T/X)$ and $g(X_S/X_T)$?

The first step is for us to escape the ubiquitous implicit assumption that all quantifiable uncertainty is due to sample selection bias. Obviously no one believes this. Researchers are aware that measurement error, selection bias, confounding, and model specification contribute to total uncertainty. But the test statistics reported in epidemiology lock in a way of thinking that leaves other sources of uncertainty unquantified.

It is difficult to attract research attention to a problem when there is no demand for the results. Doing cost-benefit analysis to figure out when more research is warranted is a tractable problem and a worthwhile endeavor. If researchers, policy makers, and funding agents recognize the value of such analysis, the demand will induce the necessary methodology research. The result could be a huge boost in the efficiency of the field of epidemiology at a relatively modest cost. More research is . . . likely to be worth its cost.