

Further Beyond the Confidence Interval

Carl V Phillips, MPP PhD

University of Texas School of Public Health and UT Medical School

and

Center for Philosophy, Health, and Policy Sciences, Inc.

presentation at the 36th Annual Society for Epidemiologic Research meeting

Atlanta, Georgia

12 June 2003

Point estimates should not be reported.

Point estimates should not be reported.

To expand a bit,

for research reports of exposure-disease relationships,

the point estimate typically does not provide genuinely useful information,

and reporting it creates several problems.

Quantifying uncertainty in health research, an oversimplified history of the last 50+ years

reasoning from way-too-small studies

frequentist hypothesis testing

(unfortunate detour into just reporting S / NS)

p-values

confidence intervals

(minimal detour into p-value functions)

(parallel development of Bayesian methods - likely more useful, but less well known)

History, continued

Quantifying multiple sources of uncertainty
quantifying random errors is not enough!

Mechanically a close relative of sensitivity analysis, but SA asks "what happens to the point estimate if we are wrong about....?"

Quantifying uncertainty recognizes "we know we are uncertain about....," and optimal decision making requires us to quantify and report that uncertainty.

recent work by Greenland, Lash, Maldonado, Phillips, et al.

Take home point: you can't make good policy decisions without this.

TALKING NOTES FOR PREVIOUS SLIDE:

Five years ago, having come to epidemiology from policy analysis, I started asking how epidemiology analyzed the magnitude and sources of uncertainty, since you need such information for making good policy decisions. The answer I got was "we don't know how to do that".

Four years ago at SER I suggested a few methods for quantifying multiple sources of uncertainty in health research. More important, in my mind, I also tried to explain **why** we should move further beyond the point estimate, beyond the confidence interval, and beyond just doing sensitivity analysis.

The widespread support for this idea was evidenced by....: my grant applications being repeatedly turned down because this supposedly could not be done, and my papers repeatedly rejected for reasons that included the claim that quantifying uncertainty might let the world know that our results are not as certain as we would like to portray them. But the good news is that despite those problems I had, the people presenting in this session and others have made impressive progress in developing methods (see previous papers by each of us and my new paper at BMC Medical Research Methodology and my and Tim's papers in the next Epidemiology).

But, as I said, my message was not so much about particular statistical and numerical methods as it was about the entire purpose and use of quantitative health research. Now that we have shown that uncertainty distributions can be calculated, it is time to take the next step, move beyond treating intervals or distributions as if they are some ancillary addition to the point estimate (the "real answer"), and move them to being The Answer – the real result of a study.

So, where does the point estimate fit into the quantified uncertainty and the decision-making question?

So, where does the point estimate fit into the quantified uncertainty and the decision- making question?

It doesn't.

We need to calculate expected values of outcomes, probabilities that a value is in a certain range, etc.

I never report a point estimate when quantifying uncertainty – it just doesn't make any sense.

The distribution is the message

To borrow from Gould (and probably many others), the Median is *not* the Message.

(More precisely for most of our calculations, the geometric mean is not the message, but it does not have the same ring to it. I started with the catchier "the interval is the message," but this is not quite right.)

Why is the point estimate not the message?

For "which of two treatments is more likely to work?",
the (error corrected!) point estimate is sufficient.

For most decisions that are more complicated, it is not.

Why is the point estimate not the message?

For "which of two treatments is more likely to work?",
the (error corrected!) point estimate is sufficient.

For most decisions that are more complicated, it is not.

Most decisions are more complicated.

Why is the point estimate not the message?

For "which of two treatments is more likely to work?",
the (error corrected!) point estimate is sufficient.

For most decisions that are more complicated, it is not.

Most decisions are more complicated.

Need to trade off probability-weighted absolute risks and benefits.

So, need the probabilities of particular values.

Yes, it is hard, but decisions are made based on our research. Really.

If you want to be useful, you have to provide the information needed.

Too hard? Imagine trying to figure out the right exercise regimen for astronauts on the Space Station. NASA has to decide something.

The Distribution is the Message: A Policy Practicum in One Act

(The play's the thing. Wherein I'll catch the conscience of the epidemiologists.)

Characters:

Research Scientist

Policy Maker

Chorus (providing a running policy analysis calculation)

Falstaff

Ghost

Setting:

Regulatory hearing to determine the future of a product that has been linked to a cancer.

RS: We found a relative risk of 1.9. You should ban this exposure.

PM: First, I don't see how you presume to make a policy recommendation without considering the *absolute* risk, let alone the cost of the policy – there is value in this product you want to ban. But since that is not the topic of today's lesson ...er... hearing, I will let you off the hook and concentrate on uncertainty.

Chorus (*shuffling through a stack of papers*):

If the research is exactly right...

Let's see, based on exposure prevalence and absolute risk, banning the exposure would eliminate 27 fatal cancers/year (a cheaper regulation would eliminate about half that).

If the research is not exactly right... (*shrugs*)

Obviously the research is not exactly right, and no doubt our wise policy maker knows that. Out, damned point (estimate)! out I say!

PM: You can't say the risk is elevated by exactly 90%. Tell me about your uncertainty.

RS: I'm impressed – I didn't know you people actually understood about uncertainty.

The true value might be anything across a wider range, but our confidence interval is (1.0001, 3.6), so we rule out the possibility that there is zero increase in risk. Therefore we should ban the exposure.

Chorus:

If the true value is one of the extreme values for the CI, a ban would save 0 or 78 lives.

But there are more things in heaven and earth, Horatio, than are dreamt of in your confidence interval calculation... The value is not one of those extreme values. Once more, into the breach.

PM: You have given me three ratio values. The probability that any one of them is exactly right is zero. How am I supposed to make a decision based on that?

And that does not even account for possible biases. I am going to have to deal with stakeholders who claim that your research has biases and should be ignored, and others who will point out that a null result is compatible with your data.

PM: You have given me three ratio values. The probability that any one of them is exactly right is zero. How am I supposed to make a decision based on that?

And that does not even account for possible biases. I am going to have to deal with stakeholders who claim that your research has biases and should be ignored, and others who will point out that a null result is compatible with your data.

Chorus: The fault, dear Brutus, is not in our standard errors but in ourselves.
Or maybe it's in our non-standard errors.

PM: Who let this guy in here? Don't we have any security in this building?

A brief scuffle ensues.

RS: I'm really impressed – who told you that our studies have biases?

Actually, now that you know our secret about the biases, I have been authorized to tell you that we believe that measurement error has biased our result downward. We calculated the resulting uncertainty intervals based some recent articles from *Epidemiology* and *Epidemiologic Perspectives & Innovations*.

RS: I'm really impressed – who told you that our studies have biases?

Actually, now that you know our secret about the biases, I have been authorized to tell you that we believe that measurement error has biased our result downward. We calculated the resulting uncertainty intervals based some recent articles from *Epidemiology* and *Epidemiologic Perspectives & Innovations*.

PLUG FOR NEW JOURNAL

Please pick up a flyer about *Epidemiologic Perspectives & Innovations (EPI)*, a new (coming soon) online peer-reviewed journal that will emphasize methods, policy, ethics, teaching, and other underrepresented aspects of health research.

RS: The result is that we believe with 90% certainty that the true, bias corrected RR is between 1.5 and 3.5, with a median value of 2.3.

PM: Now I'm impressed. I didn't know you people kept up with current methodology. So, despite your annoying habit of using relative ratio measures, which makes my analysis more difficult, I can work with that.

Chorus:

Approximating the mean value based on those numbers, yields an expected value of appx 45 fatal cancer/yr.

We all know, of course, that an expected value for absolute risk is the minimum amount of information for actually doing a policy analysis. We only now have enough information to calculate it.

With that I, your humble chorus, will now retire to crunch some numbers.
(exits)

PM: Finally! Enough information to justify a ban – if I were really facing an either-or question. But I actually need to figure out whether it would be better to do less drastic regulation, call for more research, or something else. I can do that with enough information. Can you give me a complete probability distribution of true values, dose-response, and a breakdown of what the sources of uncertainty are.

RS: You are in luck. My graduate student tells me I can do that . . .

PM: Finally! Enough information to justify a ban – if I were really facing an either-or question. But I actually need to figure out whether it would be better to do less drastic regulation, call for more research, or something else. I can do that with enough information. Can you give me a complete probability distribution of true values, dose-response, and a breakdown of what the sources of uncertainty are.

RS: You are in luck. My graduate student tells me I can do that . . .

... if you will just give us another \$20,000 grant to figure it out. You see, we already used up our entire half million dollar budget on the field study and so our data analysis consisted of a student running the results through Stata as part of her thesis. Who knew that we would have to work to analyze the data. I thought it would just speak for itself.

Exit Research Scientist

Enter Falstaff and the Ghost of Speaking Data

Falstaff: If you are reading this, Carl is not advancing the slides quickly enough

Ghost of Speaking Data: If you are reading this, Carl is not advancing the slides quickly enough.

Falstaff: If you are reading this, Carl is not advancing the slides quickly enough

Ghost of Speaking Data: If you are reading this, Carl is not advancing the slides quickly enough

Everyone falls dead from poison.

NOTE TO READERS WHO DID NOT SEE THE TALK: As you might guess, the game was to flash past this fast enough that the audience could only catch the bold part and not see the text. (I guess y' kinda had to be there to appreciate this and the performance in general.)

Alas, the tyranny of the stopwatch leaves us no time for clowns and ghosts.

It was a tormented journey to get to useful information

If forced to make a decision based on the point estimate (with or without the confidence interval), our PM would face an inadequately informed policy "debate":

"It could kill hundreds of people over the next decade. Ban it!"

"It could be harmless. It is all much ado about nothing."

"There is a lot of uncertainty. We should wait until we know more."

"There is a lot of uncertainty. We should ban it until we know more."

The first sentence of each of these is true.

Sufficient information for decisions

A best estimate that there is harm is not sufficient to justify action.

"Proving" there is harm (that a null relationship is not compatible with the data) is not sufficient.

Being uncertain is inevitable, so not sufficient for anything.

Yet, the "debate" will happen. Will it carry the day?

If we arm policy makers with an expected value (i.e., average, considering all uncertainty) or – much better and not much harder – a complete distribution of true values, then science has a chance of carrying the day.

It is no guarantee, but failing that, it is inevitable that science will not carry the day.

Other reasons why we need to calculate and report uncertainty distributions:

We cannot correctly combine multiple findings about the same E-D relationship without fully quantified uncertainty.

We cannot figure out the value of further research without an estimate of where current uncertainty stems from (more research is needed?).

(For more on probability distributions and why you cannot make a policy decision without them, see my 2001 IJE paper or a basic text on policy analysis)

But why *abolish* the point estimate (and, more so, dichotomous statements of findings)?

(1) The public fixates on simple answers.

<insert your own joke about the Bush administration or weight loss diets>

This is also true for most policy makers, who may not measure up to the character in our play.

Give a simple answer, and they think they have learned something.

Witness the half minute in every half hour news show spent reporting stock market index values.

Nice clean numbers.

Of absolutely no value to any listener who does not already know them.

Worse than worthless, they give an illusion of learning about the economy.

Reasons to abolish, cont.

(1.a) When someone fixates on a simple answer, they get very unhappy when it changes.

We tell them one thing (with an air of certainty) one week, and something else the next week.

Stop telling them things with an air of certainty and this problem goes away.

They will either have to learn more or realize they do not know enough.

Reasons to abolish, cont.

(2) It is much easier to manipulate a reported point estimate (and probably, more so, a confidence interval) than a complete probability distribution.

(for some examples of manipulation strategies, see my posters later today)

Reasons to abolish, cont.

(3) Researchers who report a point estimate think they have provided enough useful information.

If they are discouraged from reporting the point estimate, they will have to consider how to actually provide useful information.

(3.a) When researchers *calculate* a point estimate, they fixate on it.

Random data exercise (Poole; Goodman & Phillips)

- Give a class of epi students datasets drawn from a single population,
- Let them analyze the data to get relative risks and confidence intervals,
- Interesting results (for 5%, the true value really is outside 95% CI)
- (more details on this teaching strategy will be available soon)

ANOTHER PLUG FOR THE NEW JOURNAL

Epidemiologic Perspectives & Innovations will fill a void by publishing articles about teaching methods and other strategies for improving understanding of epidemiology. We hope that more information about this example will be one of the first such articles.

ANOTHER PLUG FOR THE NEW JOURNAL

Epidemiologic Perspectives & Innovations will fill a void by publishing articles about teaching methods and other strategies for improving understanding of epidemiology. We hope that more information about this example will be one of the first such articles.

Key: Students dutifully calculate confidence intervals, but they still "know" that their point estimate is right

The only apparent solution: Lose the point estimate.

Conclusions

Chorus:

A brighter day dreamt of, where applied researchers provide research that actually can be applied,

where press releases and abstracts contain more than sound and fury, signifying nothing.

Good night, good night! Parting is such sweet...

Exit Chorus, dragged off by a security guard, for pushing one quotation too far.

Exit The Point Estimate