

Debate

Open Access

Quantifying errors without random sampling

Carl V Phillips*¹ and Luwanna M LaPole²

Address: ¹Management and Policy Sciences, University of Texas School of Public Health and Center for Clinical Research and Evidence Based Medicine, University of Texas Medical School Houston, Texas, USA and ²University of Minnesota School of Public Health, Minneapolis, Minnesota, USA

Email: Carl V Phillips* - cphillips@sph.uth.tmc.edu; Luwanna M LaPole - lapo0004@tc.umn.edu

* Corresponding author

Published: 12 June 2003

Received: 16 July 2002

BMC Medical Research Methodology 2003, **3**:9

Accepted: 12 June 2003

This article is available from: <http://www.biomedcentral.com/1471-2288/3/9>

© 2003 Phillips and LaPole; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: All quantifications of mortality, morbidity, and other health measures involve numerous sources of error. The routine quantification of random sampling error makes it easy to forget that other sources of error can and should be quantified. When a quantification does not involve sampling, error is almost never quantified and results are often reported in ways that dramatically overstate their precision.

Discussion: We argue that the precision implicit in typical reporting is problematic and sketch methods for quantifying the various sources of error, building up from simple examples that can be solved analytically to more complex cases. There are straightforward ways to partially quantify the uncertainty surrounding a parameter that is not characterized by random sampling, such as limiting reported significant figures. We present simple methods for doing such quantifications, and for incorporating them into calculations. More complicated methods become necessary when multiple sources of uncertainty must be combined. We demonstrate that Monte Carlo simulation, using available software, can estimate the uncertainty resulting from complicated calculations with many sources of uncertainty. We apply the method to the current estimate of the annual incidence of foodborne illness in the United States.

Summary: Quantifying uncertainty from systematic errors is practical. Reporting this uncertainty would more honestly represent study results, help show the probability that estimated values fall within some critical range, and facilitate better targeting of further research.

Background

Most health statistics are reported with an explicit quantification of uncertainty because they are based on a sample from a target population (possibly with random assignment of treatments), and quantifying the resulting stochastic error is done almost universally. Extrapolations from samples are not, however, the only way to calculate rates, totals, or other quantitative measures of health. The availability of data may lead to an approach that does not involve sampling or any other random process. For example:

- Automobile fatality totals are typically computed with an attempt to completely enumerate, counting every case.
- Two states might create a "natural experiment" by having different traffic or safety regulations. Differences between or ratios of frequencies of accidents or injuries could then be computed by enumeration and arithmetic.
- Samples of convenience are extrapolated to the entire population, such as trying to impute the U.S. incidence of *Escherichia coli* O157:H7 infections based on data from the few states that report good data. While this is a sample,

the error comes not from random sampling error (which would be quite small) but the other sources identified below.

- To estimate the rate of a disease in a community, we frequently reverse this process and interpolate from a national average.
- Trends may be inferred from data that is believed to be related to the measure of interest, but in unknown ways, such as tracking the effect of economic changes on mental health by tracking the number of calls to hotlines.

Observational studies and randomized trials almost always quantify error due to random sampling and allocation. The realization, which was close to universal as long as four decades ago, that health research results need to include such quantification has helped reduce the frequency of conclusions based on inadequate sample sizes. However, the ease of quantifying that single source of error has distracted researchers from the many other, often larger, errors in any estimated quantity. Quantification of the one source of error implies that this represents all uncertainty. This implication grossly overstates precision by ignoring uncorrected confounding, selection biases, measurement errors, and specification of functional relationships.[1-4] This is especially clear when a calculation does not involve sampling and, lacking the one source of error that we commonly quantify, the numbers are reported as point estimates with no acknowledgment of uncertainty at all. A complete lack of error quantification implies the even more misleading claim that the result is perfect. While some readers recognize the inevitable uncertainty and guess at its magnitude, most do not.

Two highly publicized recent examples illustrate this. The death counts from the September 11 attacks on New York were updated hourly, and reported to four significant figures (the exact count). But the reports from the first few weeks turned out to be high by a factor of two, making it quite clear that even the apparently precise counting of fatalities from a single event can only be estimated within a range of error. The vote count in Florida in the 2000 U.S. presidential election involved complete enumeration. People were shocked that there was so much uncertainty – due to measurement and recording errors, among other things – in what they imagined to be a flawless mechanistic process. Few people understood that the results from the various counts represented a statistical tie, and that choosing which vote count was the "right" one was a matter of legalistic detail rather than scientific truth. (Note that this considers only the votes as counted. The illegal disenfranchisement of tens of thousands of eligible voters – who would have almost certainly broken the tie –

reminds us that uncorrected systematic bias can have much larger magnitude than the measured result.[5])

Analysis and discussion

Recently, methods have been developed to quantify the combination of various random and systematic errors in epidemiologic studies.[1-4,6-8] Simpler versions of these methods can be used to quantify errors in estimates that do not involve sampling. The following analysis demonstrates how this can be done (and by implication, why it should be done), for an increasingly complicated set of examples.

A simple method for quantifying errors in simple statistics

A quick and easy way to avoid overstating precision is appropriate rounding, as taught to high school science students (and largely ignored in health science reports, though a brief exposition of the point can be found in a recent epidemiology textbook [[9], p.51]). This method is rough (and thus not perfectly well-defined), but it is a fairly effective shorthand: do not report significant digits (i.e., digits other than place-holder zeros) beyond the level of precision of your estimate. If your point estimate for some value is 2.3456, but you think it is fairly likely that the true value is lower or higher by as much as 5%, only report 2.3. This can be interpreted as roughly, "we are pretty sure the result is between 2.25 and 2.35, but cannot be much more precise." Similarly, if your estimate is 87,654 but you know the measurement is only precise to plus-or-minus five thousand, report 90,000.

The limits of this method are clear when you consider what to report in the first example if you want to reflect confidence of plus-or-minus 15%. Reporting 2.3 implies a bit too much precision, but reporting 2 implies too little. It usually makes sense to imply a bit too much precision rather than too little (thus providing more information about the point estimate), but we should stop at the minimum level of over-precision possible (2.3 in this case) and not imply more precision still (e.g., by reporting 2.35).

Annual U.S. automobile accident fatalities are reported to five figures (e.g., 41,611 for 1999 [10]), but when presenting this result for most purposes, it is better to report 42,000, roughly estimating the limitations of measurement (e.g., some deaths should be counted as suicides or were fatal cardiovascular events before the crash) and record keeping (e.g., cases inadvertently recorded and reported by two different jurisdictions).

Notwithstanding the lack of a perfect rule, it should be clear when a result is presented with far too much precision, as is often the case. One of the most influential epidemiologic publications of recent years, Kernan et al.'s

[11] study of phenylpropanolamine and stroke (that resulted in that popular decongestant and diet aid being removed from the U.S. market) reported one odds ratio of 15.92, even though one of the cell counts generating that ratio (exposed noncases) was exactly 1, and thus it could only be precise to about 1 part in 2, not 1 in 1000. (Consider that if the cell count differed by the minimum possible, 1, the odds ratio would either be cut in half or increased to infinity.) It is difficult to assess exactly what impact this misleading claim of precision had on policy makers and other readers, but we might suspect that it contributed to the unwarranted confidence in the study's findings.

If a more formal quantification of uncertainty is reported – such as reporting "2.35 plus-or-minus 0.12" for the above example – then the significant digits are no longer the only reporting of uncertainty, and are not so important. Nevertheless, if 2.3456 appears in a paper, there is a risk that it will be repeated out of context in all its implicit precision, without the +/- .12 clarification.

It should be noted that rounding to an appropriate number of significant digits (or any other method in this paper) does not *create* any imprecision; the imprecision exists even if we do not accurately report it.

Improved quantification of errors in simple statistics

Eliminating implicit over-precision is an important step, but ultimately our goal should be to report our estimates as realistic ranges of true values by quantifying the sources of error. The simplest case is when all but one of the sources of uncertainty are *inconsequential*. (A source of uncertainty is considered inconsequential when it is enough smaller than other sources of uncertainty that it can be ignored. The implications of this should become clear as the concept is employed.) In such cases, a probability distribution for the one consequential uncertainty can be reported and used in any calculation in place of the point estimate.

The probability distribution for values of a quantitative measure can be estimated from validation studies, ranges of values observed across studies, or the best expert judgment of the researchers. Any of these are likely to be better than failing to quantify uncertainty, in effect saying "we are not really sure whether the uncertainty is small or large, so let's just call it zero." If we cannot even roughly estimate how big the errors might be, then we are reporting a result that may not even remotely reflect the true value. If we believe we are more certain than that, we should be able to at least roughly quantify how certain we are.[1,3] We may not be able to estimate the distribution precisely, but even by estimating it roughly we can avoid

unwarranted claims about the precision of the point estimate.

We propose that when statistics are reported, researchers should estimate uncertainties in the inputs used to generate the statistics, calculate the combined results of those uncertainties to arrive at the uncertainty of the final result, and determine how to accurately and parsimoniously report that uncertainty. The present discussion sketches a method for the second of these steps and touches on the third. The details of the methods for the first step, such as validation studies, are beyond the present scope.

Example

A report presented to the public and policy makers states that 2.76 percent of the people in a community have been diagnosed with a certain disease during a one-year period, based on active monitoring that identified 8650 people diagnosed out of a community of 312,962. Because the study method was a complete enumeration, there is no random process, and thus no frequentist error statistics. The resulting lack of a confidence interval means that no statement of error accompanied the result. It is certain, however, that there is still error. In particular, the researchers believe that 8650 is an undercount of as much as 20%, due to the likely inability of the monitoring system to detect all cases.

The total population of the community is also uncertain, but this is inconsequential (by the above definition). If reporting the figure in a final report, it would probably be appropriate to report 313,000 rather than the six figures, but this uncertainty is dwarfed by that of the numerator (on the order of 1 part in 100 or even 1 in 1000, compared to 1 in 10 for the numerator), and so can be ignored in the calculation. Even setting aside the downward bias, the precision implied by 2.76 percent – that we are fairly confident that we know the true value to about 1 part in 100 – is unwarranted.

After further contemplation and examination of validation data, the researchers decide that their best estimate is that the raw estimate is low by between 0 and 20 percent of the estimated value, uniformly distributed. The process by which they came to this conclusion – possibly involving a series of calculations based on the quality of monitoring, test sensitivity, etc. – is beyond the present scope. One might dispute the implicit claim that there is no chance of false positives, but it should be remembered that uncertainty distributions are never going to be perfect or beyond criticism. The researchers are simply of the opinion that the number of false positives will, with extremely high probability, be exceeded by the undercount.

(Rather than bundling these two sources of error into a single distribution, using intuition that may or may not be right, the researchers might have been better off reading ahead and using one of the methods for combining multiple sources of uncertainty. Had they done so, they might well have concluded that the misclassification error was indeed dwarfed by the under-reporting, and returned to a single quantification.)

The result of the researchers' uncertainty distribution is a uniform distribution for the annual disease incidence over the range [2.76%,3.32%]. How should this be reported? One option is to report 3%, which, conveniently, is the rounded result for the entire range. It accurately implies that our precision (with a known error of up to 10% on either side of the mean) warrants only about one significant figure. To provide more precision, the certainty interval containing 50% or 90% of the probability mass could be reported. (Again, without implying too much precision for the boundaries. An uncertainty distribution should itself not be stated in an overly-precise manner.) It is usually not a good idea to report the extremes and imply that the corrected value certainly falls between them. Extreme values can be misleading to the reader. They are also very sensitive to the exact input distributions used, such as in the current example, where the input distributions with zero probability beyond some range are good estimates for most of the probability mass, but they exclude extreme values that the researchers do not actually believe have zero probability.

The choice and nature of subjective uncertainty distributions

A detailed assessment of what needs to be considered in developing uncertainty distributions for inputs in this kind of analysis is beyond the present scope, but it is worth making a few comments to provide perspective and help researchers get started. The uniform distribution in the preceding example provides the easiest teaching example, but is probably not realistic. Even interpreting it as an approximation, it is unlikely that someone believes some range of values are approximately equally likely, but a value slightly outside that range is (approximately) impossible.

Typically, we have a point estimate and think the true value of the parameter is likely near it and the probability drops off as we deviate in either direction. This describes various distributions, including the normal, logistic, triangular (where the probability density is unimodal, dropping off linearly to zero in each direction, forming a triangular density function), and others. The choice among distribution shapes can be made largely based on whether the researcher wants to allow values to trail off to infinity or not and whether the distribution is symmetri-

cal. A triangular distribution, while seldom appearing in nature, might effectively approximate someone's beliefs, and has the advantage for pedagogic purposes of allowing calculations using polynomial algebra. Normal and logistic distributions are easy to work with using numerical methods.

It turns out that the choice of the exact shape of the distribution, after the rough magnitude of uncertainty has been determined, is relatively unimportant. Estimates like those presented here are fairly stable across unimodal distribution shapes, as long as the probability mass is substantially overlapping. (I.e., if two distributions have a very similar range for the middle 50% of their probability mass and also for the middle 90%, they will have very similar implications in these uncertainty calculations.) It should be remembered that the purpose of these methods is to better represent uncertainty, and that goal is not well served by claiming too much precision about the details of the inputs and calculations.

The question of whether an input distribution corresponds to something found in nature brings up a complicated philosophy-of-statistics question: What exactly are these probabilities? To give an abbreviated answer, they are subjective probabilities that take into consideration all of the researcher's knowledge, except the point estimate for the parameter of interest they have calculated (in the above example that would be the annual disease incidence) and any prior beliefs about what the true value of that parameter is. The subjectivity of this probability should not be seen as surprising or regarded as a limitation. All of scientific inquiry, from hypothesis generation to study design to drawing conclusions, is a highly subjective and sociologic process. Furthermore, the alternative to specifying such a distribution is to produce calculations based on the assumption that there is zero uncertainty, which is either a subjective belief itself or (more likely) is strongly believed to be wrong.

The restriction that prior beliefs about the true value should be excluded from the researchers' generation of the input probabilities is a subtlety that relates to how we can interpret the results and how the resulting uncertainty would relate to random error if it were included in the calculations. While it is not necessary to delve deeply into Bayesian analysis to do the simple calculations proposed in this paper (or to generally revise our thinking about how certain most results are), a formal interpretation of the quantified uncertainty of a result is that it is a Bayesian posterior distribution based on a prior distribution (i.e., belief about the distribution before the research in question) that assigns equal likelihood to all values in the relevant range. To understand the importance of the prior distribution, consider the possibility that one of the

researchers in the previous example was very confident that the actual incidence of the disease was above 3%. In that case, upon seeing the results of the study, she would not believe that the whole range was equally likely, but instead would think the upper end of it was more likely because her new beliefs would be based on a combination of the study result and what she knew before.

The implicit assumption that all possible values were equally likely (called a "flat prior") is problematic, because a flat distribution across all possible values is never realistic. The next step in improving these calculations should be to relax that assumption. However, the problems inherent in the implicit assumption of a flat prior (which makes the calculations much easier to perform and to understand) are reduced by a few factors. First, the "relevant range" condition says that the prior only needs to be flat across the range that contains most of the probability mass resulting from the calculation (e.g., in the above example, it would only have to be flat across [2.76%,3.32%]). This means that the worst problem of a totally flat prior, that unrealistically extreme values have to be considered to be as likely as realistic values, is absent. Furthermore, the intuition we as readers have learned from years of interpreting point estimates and frequentist confidence intervals is to, roughly, treat them as calculations based on flat priors and then roughly incorporate them into whatever actual prior belief we have. That is, if a study reports an estimate of 3 or some interval around it, and we were previously quite sure that the true value was 5 or more, the new evidence might push our beliefs downward, but it is not going to replace them. This is just as true if the interval is a standard confidence interval (based only on random sampling error) or an uncertainty quantification, and our practiced intuition will be useful until these methods are advanced.

Increasing complexity

As the calculation of the point estimate becomes more complicated, so does the uncertainty calculation. Multiple sources of uncertainty need to be combined in the same way as their underlying point estimates – additively, multiplicatively, or otherwise – as demonstrated in a further example.

Example, continued

The researchers wish to extrapolate the frequency of disease from the study community to estimate the total cases for the entire state, with a population of 10,456,000. A naive way to introduce quantified uncertainty into the calculation would be to treat the original study of 312,962 people as a random sample from a population of 10,456,000. The result could be quantified using the usual frequentist statistical methods, with the result misleadingly suggesting high precision (a 95% confidence inter-

val of (2.71,2.82)). But greater uncertainty is introduced by the extrapolation of the results, which introduces unknown levels of non-stochastic error. Perhaps the sample community was studied because it was particularly convenient, because it was more urban or had a better local health department, and so is different from the state average.

The researchers do not know the specific amount of bias for their estimate, but they recognize that there is likely some bias. Their best estimate is that the actual rate for the state is most likely within +/-10% of the sample community, but it is plausible that the extrapolation is off by as much as 25%. To fit these probabilities, the researchers use a symmetrical triangular distribution from .75 to 1.25 of the point estimate (i.e., zero probability density at .75, linearly increasing to the midpoint, 1.0, and linearly decreasing to zero at 1.25). They could have chosen a normal distribution or various other similarly-shaped distributions to represent these beliefs with similar outcomes.

This new source of error now combines with the original underestimate to produce a probability density for the total number of cases in the state. The additional uncertainty from random sampling error is small and can be ignored as inconsequential. Alternatively, the random sampling error could be incorporated into the researchers' subjective probability. (Objectively determinable stochastic processes can be brought into the uncertainty calculation differently from subjective uncertainty, but this introduces complexity that is left for future analyses. The purpose of the present analysis is to consider cases where sampling error is absent or is insignificant compared to other sources of error.)

The density for a given final value, x, which results from a calculation involving two uncertain values is the integral across values of the two functions that produce x. In the present case, this is relatively simple to calculate. The probability distribution for the total number of cases in the state is described by the continuous approximation:

$$f(x) = k \cdot \int_{.75}^{1.25} g(t)h(s)dt$$

$$where\ g(t) = \begin{cases} (t - .75) \cdot .16 & \text{for } t \in [.75, 1.0] \\ (1.25 - t) \cdot .16 & \text{for } t \in (1.0, 1.25) \\ 0 & \text{otherwise} \end{cases}$$

$$s = \frac{x}{t} \cdot k$$

$$k = \frac{312,962}{8650} \frac{1}{10,456,000}$$

$$h(s) = \begin{cases} 5 & \text{for } S \in [1.00, 1.20] \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where the definition of $g(t)$ describes the triangular distribution and $h(s)$ the uniform distribution, and k is the scale factor to make $f(x)$ a probability density function. The continuous approximation is necessary not just for computational convenience, but because the form of the error distributions was continuous. Since our practical interest is for ranges of values, and not the exact probability for a given value, nothing is lost by this approximation.

For this distribution, the middle 90% of the probability mass falls in the range 260,000 to 380,000. A normal distribution with a mean of 1.0 and a standard deviation of .11 would also have represented the researchers' beliefs about the bias from extrapolation, and would have yielded the same interval (after rounding) for the middle 90% of the probability mass.

Solving this equation (and thus figuring out uncertainty intervals) is easy. But adding much more complication makes it unwieldy. A third layer of multiplicative uncertainty would require a double integral over a more complicated product, and so on. An uncertain input that entered the equation other than by multiplying would be more complicated still. Indeed, simply using the normal distribution for the uncertainty from the extrapolation would make this calculation considerably more complicated. The implication is clear: with more than a few simple sources of uncertainty, closed-form (analytic) calculation is not a practical method for quantifying it.

Estimating complex combinations of uncertainty

Any calculation with a large number of inputs is likely to resist closed-form calculation of uncertainty and intuitive statements about total uncertainty are likely to be worthless. ("Large," in this case, can mean as few as three or four inputs if they all introduce uncertainty.) However, there are tools developed in finance and engineering that can be used to calculate uncertainty in such health research.

To estimate the probability density for parameters of interest given multiple uncertain input values, we propose using Monte Carlo (random number-based) numerical methods as follows:

1. Probability distributions are specified for the inputs, as presented above.
2. A random draw is made from each of those distributions, producing one set of possible true values. The calculation (the same one used to generate the point estimate) is carried out for those values to produce a possible final value of the estimate.
3. Step 2 is iterated a large number of times, producing a new candidate value for each new set of random draws.

4. These values are recorded, and can then be used to calculate the probability of the true value being in a particular interval or grouped into fixed-width intervals and represented by a histogram that approximates the probability density function for the total uncertainty.

This approach takes a difficult problem and approximates the answer by carrying out simple calculations a large number of times. Since these simple calculations are the ones that had to be constructed to get the point estimate in the first place, a relatively small amount of extra effort is required. Monte Carlo simulations of this sort are used extensively for similar calculations of uncertainty in business and engineering applications (often under the rubric "risk analysis"), and so there is user-friendly off-the-shelf software that does these calculations. (Further background in these applications is available, at the time of writing, from the manufacturer of the software we used at <http://www.crystalball.com/risk-analysis-start.html>.)

Extremely complicated Monte Carlo simulations are used to model everything from nuclear explosions to biological evolution, but the tools needed for present purposes are usable by any competent quantitative researcher.

Monte Carlo uncertainty calculations have been proposed for the errors in a typical epidemiologic study, [1-4,6,7] which are much more complicated than the errors considered in the examples presented here. Such applications involve complicated interactions of random error, selection bias, measurement error, and other sources of uncertainty that compound in mathematically complicated ways. For the straightforward adding and multiplying used in the examples presented here, these calculations are simple to program and do not require much computer time.

Indeed, the biggest challenge for quantifying uncertainty in these calculations – quantifying the various input uncertainties – is partially ameliorated by the ease with which different values can be run to produce sensitivity analyses. While sensitivity analyses cannot determine which values are better to use, they can point out which ones matter enough to warrant more attention.

A Monte Carlo-based uncertainty calculation

As an example of a calculation combining many sources of uncertainty, we use a simplified version of Mead et al.'s frequently-quoted calculation of the incidence of food-borne disease in the U.S.[12] (The present example, a highly simplified version of the second author's master thesis, [13] is intended primarily to illustrate the method rather than explore the details of food safety data. Powell, Ebel, and Schlosser have also conducted a Monte-Carlo-based analysis of the uncertainty in some of Mead et al.'s numbers.[14] An in-depth analysis of the uncertainty for

a particular foodborne disease risk can be found in Vose et al.[15])

Mead et al.'s calculation was based on literally hundreds of input numbers, including:

- the U.S. rate of total gastrointestinal illness,
- foodborne illness case counts from passive surveillance (data reported voluntarily from clinics) and outbreak reports (health department investigations),
- case counts from active surveillance (attempts to enumerate every case) at five county- or state-level sites,
- several estimates of how many unobserved cases are represented by one observed case (in order to extrapolate from data to true values),
- estimates of the fraction of cases of each disease that should be attributed to food,
- several other extrapolations.

The need to understand and quantify uncertainty is obvious when we observe that Mead et al., despite the clearly uncertain data and calculations, emphasized (in the abstract and press releases) a result to two significant figures, 76 million U.S. cases per year, with no statement of uncertainty or a plausible range of values. Widely quoted in this form, this number implies that the experts are confident that the result is not 75 million or 77 million, but is somewhere between 75.5 and 76.5 million. After all, they did not report 80 million, which would tend to imply less precision. (Note a weakness of relying on significant figures alone: If they had reported 75 million, it would not have been clear whether they were rounding to the nearest 1, 5, or 25 million. This would only become clear if multiple numbers, all with the same rounding, were reported.) The body of the Mead et al. paper actually contains an estimate to ten significant figures. (Such overly precise reporting could be justified in a paper if intended to help the reader duplicate the calculations, rather than as a conclusory statement. This is unlikely to be the explanation in the present case, since their calculation is difficult to duplicate from the information given and this kind of replication is not common in health research.)

Monte Carlo-based numerical methods allow us to estimate the uncertainty for calculations as complicated as Mead et al.'s. The complexity of their effort to use existing, highly incomplete data to estimate U.S. foodborne disease incidence is illustrated by the spreadsheet we developed to duplicate their calculations, which includes over 200 numerical cells, more than 50 of which are inputs

from outside information. Even if we believe that the analysis reflected science's best knowledge about these values, we can be sure that such an estimate is not accurate to better than 2%. But what more can we say about the range of possible values?

To answer this, we examined the various inputs and the certainty of their sources, and developed a model that incorporated estimates for each source of uncertainty. For the current example, we use a simplified version of the calculation, reducing the list of 28 different diseases to the 3 that contributed the most to the total plus an "other" category, and simplifying some of the multiplicative correction factors used in the original. The first example presented here uses conservative uncertainty distributions that are relatively small and mean-preserving (i.e., they use the original Mead et al. point estimates as the distribution mean). Even with this optimistic level of uncertainty, it is easy to see the need to avoid implying too much precision.

The calculation is summarized in Table 1. It starts with incidence rates of several diseases that are partially attributable to foodborne transmission. Two of the three major contributors are based on incomplete samples from monitoring efforts. They are multiplied by a factor of 38 to estimate the total incidence. The total incidence of each disease is then multiplied by an estimate of the portion of cases that are foodborne. These figures are summed across the diseases to get the incidence of foodborne illnesses attributable to known pathogens. The total cases from unknown sources is then calculated by estimating the total cases of gastroenteritis in the U.S. and subtracting the cases of known origins. To get the portion of these that are attributable to foodborne pathogens, Mead et al. assumed that the foodborne portion is the same as that for known pathogens. This result is added to the incidence from known pathogens to get a total.

Every one of these inputs introduces uncertainty. To reflect this, we introduced the following distributions. (Most of these distributions have some external basis. However, this calculation should be seen primarily as a demonstration of the methods for doing the analysis and a rough estimate of the minimal uncertainty in Mead et al.'s number, rather than a claim that these are the right distributions.) For the total cases of each of the three identified diseases, the point estimate is replaced by a normal distribution, with a mean of the point estimate and a standard deviation of 10 percent of the point estimate. For the 25 other pathogens, which each contributed a relatively small portion of the total, we simply used the point estimates because the uncertainty for each is inconsequential. (Assuming their errors are uncorrelated, they tend to average out, leaving a relatively tight distribution of total

Table 1: Simplified calculation of foodborne disease incidence.

Infectious Agent	Estimated Total Cases	Observed Cases	Estimated % Foodborne Transmission	Foodborne Incidence
Campylobacter spp	2,453,926	64,577	80	1,963,141
Salmonella, nontyphoidal	1,412,498	37,171	95	1,341,873
Norwalk-like viruses	23,000,000		40	9,200,000
Other	11,763,217			1,309,910
Total for Known Pathogens	38,629,641			13,814,924
Multiplier, observed to total cases		38.0		
Total cases of gastroenteritis			210,813,750	
of unknown origin (subtract)			172,184,109	
Illness of known etiology, % foodborne (divide)			36%	
Foodborne illness, unknown etiology (multiply)			61,986,279	
Total foodborne illness (not rounded)			75,801,203	

Boldface = parameters with uncertainty distributions in the example. *Italics* = input parameters (other numbers are calculated within model) An downloadable interactive version of this calculation, which can be used to run the Monte Carlo simulation to estimate total uncertainty, can be accessed via a link in the text.

uncertainty. If one believes that inaccuracies in the estimates of incidence rates are correlated across diseases, the overall uncertainty would be greater.) For the multiplicative factor used for two of those, we used a symmetrical triangular distribution centered on the original 38, with a range of (24,52).

For the percent of each disease attributable to food, we used the point estimates for two of the pathogens because they were fairly close to 100% (leaving little room for error) and appeared to be reasonably solid estimates. The remaining pathogen, Norwalk-like viruses, accounts for most of the total cases and the fraction that are foodborne is highly uncertain and far from either 0 or 100%, leaving room for substantial variation. Not only does this percentage affect the estimated number of foodborne cases of that disease, but it dominates the overall estimated percentage of gastroenteritis cases attributed to food, and thus the estimate for cases of unknown etiology. Given the large impact of uncertainty in this input parameter, we conduct a sensitivity analysis for its impact below. For the initial example, we modeled the uncertainty as a uniform distribution on [20%,60%], centered on the Mead et al. point estimate of 40 percent.

For total cases of gastroenteritis, we used a normal distribution with a mean of the original point estimate, and a standard deviation of 20 percent of the original estimate. To represent the uncertainty of the assumption that the portion of gastroenteritis cases of unknown origin attributable to foodborne pathogens is the same as for cases of known origin, we draw the portion of unknown cases from a normal distribution around the portion of known

cases (after it is calculated, since it varies with each iteration) with a standard deviation of 8 percentage points.

This example was constructed and calculated using a Microsoft Excel spreadsheet and the off-the-shelf Monte Carlo simulation package, Crystal Ball (Decisioneering Inc., Denver, Colorado). We ran half a million iterations of the model, producing the histogram in Figure 1 that approximates the probability density for the total number of cases that results from these input uncertainties. It would overstate the quality of our estimates to interpret this as providing precise probability estimates. But the rough cut at estimating the total uncertainty is very informative. Even with these relatively conservative estimates of uncertainty, chances are about half that the real total is outside the range of 50 million to 100 million.

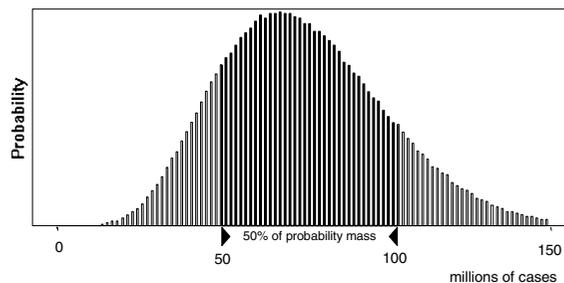


Figure 1 Approximate distribution of true foodborne disease incidence (base example)

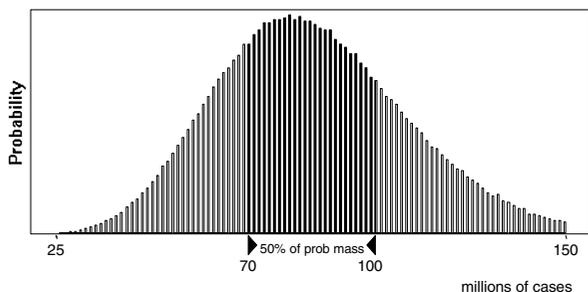


Figure 2
Approximate distribution of true foodborne disease incidence (sensitivity analysis example)

(The Microsoft Excel 2000 spreadsheet used to run the Crystal Ball simulation is available as a 1 to this paper. Running the simulations directly from the spreadsheet requires a copy of Crystal Ball 2000 or a later compatible version.)

What should we make of such a result? That depends on our goal for the estimate in the first place. If the goal is to get an estimate into the scientific literature for others to use, it is probably a good idea to report the entire distribution, along with sensitivity analyses, and let others use them as they will. Researchers interested in combining this with other estimates of the value in question might want to look at how likely the other estimates are according to this calculation and how likely this estimate is according to other calculations. A sophisticated policy maker trying to figure out how much attention to devote to this problem might want to look at the probability that the total was greater than some critical level, say ten million or one hundred million. Indeed, even a rough cut might be sufficient to answer important policy questions (e.g., "we are pretty sure that there are more than fifty million cases per year, which means this is a bigger problem than most people think").

Sensitivity analysis

This method for quantifying uncertainty lends itself easily to sensitivity analysis of the inputs. The quantification of uncertainty is itself is sometimes called a sensitivity analysis, but despite involving some of the same math, uncertainty quantification and sensitivity analysis are fundamentally different. A sensitivity analysis asks questions like, "if we are wrong about a certain input, how much does our best estimate change?" An uncertainty distribution does not answer that question. Rather, the uncertainty distribution *is* that best estimate, the best esti-

mate of the probabilities of possible true values given our knowledge. An uncertainty distribution does not report deviations from the result of an analysis; it is *the result of an analysis*. A sensitivity analysis can be done to see how much that distribution changes if we change one of our inputs.

For example, compare a second estimate, identical except that the distribution of foodborne attribution for Norwalk-like viruses is [40%,60%] instead of the previous values of [20%,60%]. (This is chosen primarily to be illustrative and emphasize the effect of changing the mean and variance of the range. It turns out, though, that Mead et al. based their input of 40% on a single rough estimate in the literature which was 47%, so the new mean of 50% is actually closer.) The result, represented in Figure 2, shows that the new mean value is about 85 million and that half the probability mass is now in the narrower range of 70 to 100 million. The substantial difference between this and our previous distribution makes clear that one of the most useful things that could be done to improve our estimate of the total cases is to reduce our uncertainty about this key input. Furthermore, it calls to mind the question of whether even experts who see the estimate of 76 million (which appears, usually presented as fact, in almost everything written about foodborne illness in the U.S.) have any idea that it hinges so significantly on this one, rather arcane, guesstimate about how much of the illness caused by one pathogen is attributable to foodborne transmission.

Summary

It is possible to quantify uncertainty in complex calculations in health research, as is commonly done for non-sample-based calculations in business or engineering. In addition to simply being a more accurate presentation of scientific knowledge, such quantification could dramatically increase the value of the underlying estimates in several ways. It would clarify whether the estimates are certain enough for the purposes for which they are used. Furthermore, it would suggest how likely further research is to produce a substantially different answer and would direct such research toward improving the particular inputs that create more of the uncertainty. The notion of reporting uncertainty sometimes provokes opposition, as if the revelation of uncertainty were responsible for creating the uncertainty. But quantification does not introduce uncertainty that did not previously exist, but rather, replaces ignorance about the magnitude of that uncertainty with the best available knowledge.

Competing interests

Phillips and a few of his students have received free short-term licences for the Crystal Ball software (though this analysis was carried out using previously purchased cop-

ies). After this paper was written and reviewed, but before resubmission, Phillips was briefly retained as an expert witness by The Delaco Company in litigation related to the phenylpropanolamine study that is briefly mentioned.

Authors' contributions

CVP conceived of the goal, developed the method of analysis, created the simple examples in the manuscript, and simplified the final example for the manuscript. LML carried out the research and original analysis of the in-depth example (foodborne illness) and created the programming for the calculation. CVP composed the manuscript; LML read and approved the final manuscript.

Additional material

Additional file 1

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2288-3-9-S1.xls>]

Acknowledgments

The authors acknowledge the contributions of George Maldonado and Craig Hedberg, collaborators in related analyses, and thank Karen J Goodman, Charlie Poole, and Steven Cole for helpful comments on the manuscript.

References

- Phillips CV and Maldonado G: **Using Monte Carlo Methods to Quantify the Multiple Sources of Error in Studies. Abstract of presentation at 32nd Annual Meeting of the Society for Epidemiologic Research, Baltimore, June 1999** *American Journal of Epidemiology* 1999, **149**(11):S17.
- Phillips CV: **Applying Fully-Articulated Probability Distributions. Abstract of presentation at 33rd Annual Meeting of the Society for Epidemiologic Research, Seattle, June 2000** *American Journal of Epidemiology* 2000, **151**(11):S41.
- Phillips CV: **Quantifying and Reporting Uncertainty from Systematic Errors** *Epidemiology* 2003.
- Phillips CV: **Quantified Uncertainty and High-Cost Public Health Decisions: The Case of Phenylpropanolamine. Abstract of presentation at 35rd Annual Meeting of the Society for Epidemiologic Research, Palm Desert, June 2002** *American Journal of Epidemiology* 2002, **155**(11):S69.
- Palast G: *The Best Democracy Money Can Buy*. Plume 2002.
- Phillips CV: **The Economics of "More Research is Needed"** *International Journal of Epidemiology* 2001, **30**:771-776.
- Lash TL and Silliman RA: **A sensitivity analysis to separate bias due to confounding from bias due to predicting misclassification by a variable that does both** *Epidemiology* 2000, **11**:544-549.
- Greenland S: **Basic methods for sensitivity analysis and external adjustment** In: *Modern Epidemiology* Edited by: K Rothman, S Greenland. Philadelphia, Lippincott-Raven Publishers; 1998:343-7.
- Rothman KJ: *Epidemiology: An Introduction* Oxford University Press; 2002.
- National Highway Traffic Safety Administration, U.S. Department of Transportation: **Traffic Safety Facts, 1999 2000** [<http://www.nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSAAnn/TSFI999.pdf>].
- Kernan WN, Viscoli CM and Brass LM et al.: **Phenylpropanolamine and the Risk of Hemorrhagic Stroke** *New England Journal of Medicine* 2000, **343**(25):1826-32.
- Mead PS, Slutsker L and Dietz V et al.: **Food-Related Illness and Death in the United States** *Emerging Infectious Diseases* 1999, **5**(5):1-39.
- LaPole L: **Annual Foodborne Illness in the United States** Master's thesis, University of Minnesota School of Public Health, Minneapolis, MN 2000.
- Powell M, Ebel E and Schlosser W: **Considering Uncertainty in Comparing the Burden of Illness due to Foodborne Microbial Pathogens** *International Journal of Food Microbiology* 2001, **69**:209-215.
- Vose D, Hollinger K and Bartholomew M et al.: **The Human Health Impact of Fluoroquinolone Resistant Campylobacter Attributed to the Consumption of Chicken** *U.S. Food and Drug Administration, Center for Veterinary Medicine* 2001.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/3/9/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

