Some Mathematics for Citizens: Gaining Quantitative Literacy Through Public Policy Debate

Daniel Kaplan, Macalester College, St. Paul, MN

Everyone needs to make decisions. Some of these decisions are personal: Should I take a screening test for breast cancer or prostate cancer? Should I eat beef, given the risks due to mad cow disease? Should I buy a car with better fuel efficiency? Some of the decisions are collective: Should the government impose stronger fuel economy requirements for cars to reduce oil consumption and CO_2 emissions? What steps should be taken to limit the spread of mad cow disease? Should we clamp down on immigration to improve national security?

Framing good answers to the myriad policy questions such as these—whether personal or collective—requires knowledge and understanding. Gaining this knowledge requires literacy; citizens need to be able to read and understand the information that is available. But for many policy issues, interpretation of the information—combining information, turning the information into a decision, realizing that the available information is not adequate to support an informed decision—requires not just literacy in the conventional sense, but quantitative literacy.

To illustrate, consider a common policy issue that affects almost everyone: should people take cancer screening tests? The answer almost always given is, "Yes. You don't want to get cancer." But a literate answer requires some knowledge and understanding, much of which is quantitative. Does a screening test prevent cancer? Obviously not. The literate person knows that screen tests are for detecting cancer, not preventing it. What's not widely understood, however, is that the word "screening" refers to a test for cancer among members of a population not at particularly high risk of it. This low prevalence has important implications for the effectiveness of tests. These implications can only be understood with a basic knowledge of conditional probability.

Even the simple-sounding question of whether a screening test reduces mortality or increases survival is filled with issues that require quantitative literacy. For example, there is ample evidence that breast cancer and prostate cancer screening increases survival rates. Yet, there is an ongoing debate about whether breast cancer or prostate cancer screening tests have health benefits. Given that screening has been shown to increase survival rates, how could it be that screening has no positive effect on health?[1]

Even more difficult to assess is the balance between the costs and benefits of the tests. In the press and political debate, this assessment is often based on anecdotes: the story of a person whose life was saved by a screening test. If the debate is to move beyond the anecdotal, what are the important measures of cost and benefit? How can the costs and benefits be compared? Supposing that one calculates a rate such as dollars spent per life saved, how is one to decide whether this number is too high to be worthwhile? These are all questions that require quantitative literacy.

Many people, familiar with the style of high school- or college-level math where there is always a single, correct answer, assume that mathematics is not broadly relevant to complex policy issues. There is rarely a mathematically derived solution to a policy problem; they are generally too complicated to have a single, best solution. Of course, quantitative considerations illuminate many aspects of policy debate. Just as important, quantitative literacy can sharpen reasoning about the issues.

At Macalester College, an interdisciplinary group of faculty from economics, geography, mathematics, political science, sociology, and statistics has been trying to identify the high-level quantitative skills and knowledge that are important in debating public policy and to convey this quantitative literacy to our students, particularly to those who otherwise would have little or no contact with mathematics. We teach this literacy in its natural setting, through discussion of public policy issues. Policy debate is a setting where quantitative reasoning is important for a much broader range of students than are traditionally attracted to mathematics. Many of the same skills that are helpful in debating public policy also apply to decision-making on a personal level and in business.

The quantitative topics that we have identified include the following:

- The meaning of one quantity must often be interpreted in the context of another. Rates provide a simple means of adjustment. Dollars spent on screening tests per life saved is an example of a rate.
- Average rates and marginal rates describe different aspects of a relationship. All taxpayers ought to understand why; even though their tax rate approaches 40%, they do not pay 40% of their income in taxes.
- Direct and indirect effects may cancel each other out. (In technical language, partial derivatives and total derivatives are different things and may even have different signs.)
- There are different models of growth and decline that are used to describe change. The most important are linear (arithmetic), exponential (geometric), and limited (e.g., logistic). Applying the intuition of arithmetic growth to a reality of geometric growth produces wrong conclusions.
- Money in the future is not as valuable as the same amount of money in the present. There are accepted ways to discount for the passage of time.
- "Best" is a concept that applies only to situations where there is a single goal. When there are multiple goals, trade-offs are created. While we can elucidate and evaluate trade-offs, there is no unique best answer. But there are satisfactory alternatives to the extreme view of reducing everything to a single goal. For example, should we reduce the emission of pollutants? There are several goals here: improving the health of people, improving environmental conditions, reducing global warming, and promoting economic activity. An informed debate about pollution requires some consideration of these goals and the benefits and costs of any proposed policy. How do we measure the benefits? How do we compare them to the costs when, in general, they will be in different, incommensurable units? The quantitatively literate person should understand that techniques such as putting a dollar value on a human life are not necessarily the craven, dehumanizing acts of capitalistic bureaucrats but a reasonable attempt to assess the balance between competing and incommensurate goals.
- Finding an association between two things is not the same thing as a demonstration of cause and effect.
- Sampling allows one to characterize the entire population. To use it effectively, you need to understand how sampling leads to variability and imprecision.
- Statistical results are often portrayed in terms of "significance." This term is often misinterpreted as meaning "reliability" or "importance."
- The concept of conditional probability, particularly with respect to diagnosis and testing, should be considered. Quantitative literacy is not just knowing the vocabulary of false-positives and -negatives, sensitivity, specificity, and prevalence. It also involves knowing typical values of these in different settings. For example, anyone who donates blood should understand that even though the HIV enzyme-linked immunosorbent assay test has an accuracy of about 99.9%, a positive test in a low-risk individual does not mean that he or she is likely to have the disease. Here is another example: There is much concern at present about how to screen immigrants to prevent terrorism. To many people, ethnic profiling is a matter of common sense; to others, it is racism. From reading the newspaper, many people seem to have some approximate notion of the link between ethnicity and terrorism. The quantitatively literate person should be able to distinguish between two different conditional probabilities: the probability that a person has a certain ethnicity given that he or she is a terrorist, contrasted with the probability that the person is a terrorist given that he or she has that ethnicity. Newspaper reports suggest that the first probability is rather high. But it is the second probability that should inform policy debate on ethnic profiling. The two probabilities are dramatically different in size; confusing them leads to bad policy.

Another such list, with a focus on statistics, can be found in an article by Jessica Utts (1).

In teaching these quantitative topics, we use a case-study approach. The quantitative topics are introduced, elaborated upon, and reviewed in the context of real-world policy decisions. We draw extensively from newspapers, both to frame the policy issue and to give examples of how quantitative illiteracy can cause reasoning to go astray.

A Case Study: Mad Cow Disease

To illustrate the case-study approach, consider the problem of mad cow disease. The public first became aware of mad cow disease, technically called bovine spongiform encephalopathy (BSE), during an outbreak in Great Britain starting in the mid-1980s. The agent of transmission is the prion; the mechanism of transmission is eating infected tissue. In Britain, the epidemic of BSE was fostered by feeding cattle on the remnants of other ruminants, including cattle. BSE causes a slow wasting of neural tissue and, it is believed, can be transmitted to humans, producing variant Creutzfeldt-Jakob disease (vCJD), the lethal outcome of which may take decades to become manifest. Initially confined to Great Britain, BSE is now found in the European continent, Japan, and North America.

The question for an individual is whether beef is safe to eat. The question for government is how best to regulate the feeding, slaughter, testing, and trade of cattle and beef.

Numbers are important here. What is the prevalence of BSE and what is the probability that it is passed on to humans in the form of vCJD? This might be effectively conveyed with a single number, perhaps in this format: how many pounds of beef must typically be eaten to give a 1% chance of contracting vCJD? Such a number, regardless of its precise format, would give valuable information similar to that available regarding HIV transmission or the safety of air travel.

Estimates of this number must be imprecise. In Britain, where the BSE outbreak was roughly 1,000 times as great as in any other country, initial estimates of the eventual number of people eventually contracting vCJD ranged from 10 to 80,000. A recent study, based on tests for the vCJD prion in stored appendix and tonsil samples, estimates that 3,800 people in Great Britain are carrying the prion. (The upper bound of the 95% confidence interval is about 10,000.) Approximately 150 people have died so far from vCJD. These numbers suggest that the risk of vCJD in the United States, where there are few BSE cases, is very small. Knowing how to deal with imprecision, understanding how imprecision stems naturally from sampling, and recognizing that imprecision is not the same as ignorance, are important aspects of quantitative literacy.

Here is an example of quantitative illiteracy. A United Press International (UPI) report[2] makes a biological claim that might be reasonable: "Several studies…have found that autopsies reveal 3% to 13% of patients diagnosed with Alzheimer's or dementia actually suffered from vCJD.[3] Those numbers might sound low, but there are 4 million Alzheimer's cases and hundreds of thousands of dementia cases in the United States." The report goes on to say that the true prevalence of vCJD is not known, partly because the rate of autopsy has dropped from about 50% to less than 15%. The article quotes a researcher as saying, "If we don't do autopsies and we don't look at people's brains, we have no idea about what is the general prevalence of these kinds of infections and (whether) it is changing."

This is dangerously wrong. Even an autopsy rate of 15% is quite capable of detecting the claimed prevalence of vCJD. The quantitatively literate person should know that although a 15% sampling will miss 85% of cases, it would succeed at identifying the prevalence so long as the sampling procedure is fair. Knowing the prevalence, the number of cases can easily be estimated.

The article goes on: "At the same time autopsies have been declining, the number of deaths attributed to Alzheimer's has increased more than 50-fold since 1979, going from 857 deaths then to nearly 50,000 in 2000. Though it is unlikely the dramatic increase in Alzheimer's is due entirely to misdiagnosed vCJD cases, it 'could explain some of the increase we've seen,' [a researcher] said." The implication of the article is that BSE is associated with the increase in Alzheimer's and that we haven't noticed this because autopsies have not been done. Before writing the article, though, the quantitatively literate reporter might have checked the credibility of the link with a little worst-case analysis. If autopsies had been performed on 5% of Alzheimer deaths, and even if 1% of these were the result of vCJD, then there would be 25 vCJD cases per year in the U.S. in year 2000 identified by autopsy. To compare, here's what the Centers for Disease Control and Prevention reports (2):

> *As of December 1, 2003, a total of 153 cases of vCJD had been reported in the world: 143 from the United Kingdom, six from France, and one each from*

Canada, Ireland, Italy, and the United States (note: the Canadian, Irish, and U.S. cases were reported in persons who resided in the United Kingdom during a key exposure period of the U.K population to the BSE agent).[2]

This is not consistent with a claimed link between Alzheimer's and mad cow disease. Qualitatively, the newspaper's qualifying statement "unlikely to be due entirely" might be true, but quantitatively it is grossly misleading.

A higher level of literacy is needed to understand why an epidemic outbreak can often be stopped by lowering somewhat the transmissibility of the disease. Safeguards don't have to be perfect to be effective, as demonstrated by both mathematical models and epidemiological experience. A popular exposition of this is available in Malcolm Gladwell's *The Tipping Point* (3).

The entire idea of graded responses is missing from news stories of mad cow disease. Newspapers often define safety in terms of perfection: zero chance of contracting vCJD. Here is an op-ed piece from the *Los Angeles Times* (26 February 2004):

> *"The recent outbreak of mad cow disease led to immediate, soothing reassurances from the U.S. Department of Agriculture, resting on what sounded like hard, scientific facts. Don't worry, the official story went, we have a rigorous inspection program designed to ensure with 95 percent certainty that fewer than one in a million cattle have the disease. Doing more than that would be unnecessarily expensive because we are already, it seems, safe enough.*
>
> *Instead of resolving to find every case of mad cow disease and eradicate it from the United States, the USDA engaged in a how-much-is-too-much conversation in which it balanced the safety of our meat supply with the beef industry's bottom line.*
>
> *That cost-benefit approach is how regulatory Washington makes decisions these days, and the mad cow fiasco is the perfect example of the moral bankruptcy of the method. With 36 million cattle slaughtered annually in the United States, the 'one-in-a-million' threshold would actually allow more than one case of made cow disease every two weeks."*

Of course, different people can evaluate safety differently—individual tolerances for risk, or evaluations of the benefits brought by taking risks, vary. The relevant aspect of quantitative literacy is understanding that "non-zero does not mean big." Smallness is properly measured in terms of widely accepted risks such as eating peanut butter, food poisoning, and choking. For example, about 30 children die each year from choking on hard candy (4), considerably larger than the number of new vCJD cases that might arise each year from BSE in the U.S.

An interesting sub-issue in mad cow disease is testing. In the U.S., the government plans to test about 120,000 cattle a year out of 37 million slaughtered cattle each year. There is a particular emphasis on "downers," those cattle whose symptoms suggest an increased probability of carrying BSE. Countries such as Japan test all slaughtered cattle. Should the U.S. do this also? Editorial writers say "yes." They focus on the direct costs of testing (pennies per pound of beef) and on the potential benefits (reopening a market for the export of beef that was closed when BSE was identified in one U.S. cow).

A relevant bit of quantitative literacy here is the notion of false-positive and false-negative rates and the associated terminology of sensitivity, specificity, and prevalence. Tests are rarely perfect. The test for BSE seems to be excellent (5). The specificity is described as 100% based on a test of 6,042 samples. This suggests that there would be no false-positive results in testing cattle. But general experience with tests—including DNA fingerprinting in criminal trials—suggests that the specificity is never exactly 100%. Indeed, for basic sampling reasons, the test of 6,042 samples cannot rule out a specificity of 99.99%. If the actual specificity in practice was 99.999%, the testing of 30 million slaughtered cattle each year would produce 30 positive tests for BSE each year. It seems hard to believe that countries that have banned U.S. exported beef based on a single positive test would restart the trade during a testing regime producing tens of positive tests each year, so the proposed trade benefits of comprehensive testing are probably small.

It goes well beyond quantitative literacy to know what will be the specificity of a particular test. But the literate person should be aware of the possibility of false-positives and should know something about the rates of commonly encountered tests. The literate person should be aware that perfection in a test is hard to achieve and even harder to demonstrate.

Quantitative literacy can illuminate issues that students care about; by teaching quantitative literacy in this context, we hope that the policy issues will lead students to care about quantitative literacy. Just as verbal literacy goes beyond reading and writing to interpreting and evaluating information and placing it in a broader context, by exposing students to a quantitatively literate discussion of policy issues, we help them to carry on the literate consideration of the issues they will face in their personal and civic lives.

REFERENCES

1. Utts, J. 2003. What educated citizens should know about statistics and probability. *American Statistician* 57: 74–79.
2. http://www.cdc.gov/ncidod/diseases/cjd/cjd_fact_sheet.htm
3. Gladwell, M. 2002. *The Tipping Point: How Little Things Can Make a Big Difference.* Boston, MA: Back Bay Books.
4. http://www.cdc.gov/ncipc/duip/spotlite/choking.htm
5. http://www.idexx.com/production/ruminant/ruminant8.cfm

Endnotes

1. Apparently, screening, in addition to finding dangerous cancers, detects some cancers that would not be harmful if not detected. But these harmless cancers cannot be distinguished from harmful ones, so the person with the harmless cancer becomes a patient whose survival is credited to screening. Even if post-screening treatment saves not a single life, thanks to screening, there are more patients and hence fewer deaths per patient. Screening also identifies fatal cancers at an earlier stage, leading to longer survival times until death—which might have come at the same time even if there were no screening.
2. Human mad cow cases may already be at epidemic levels in the U.S. (21 July 2003).
3. Although the UPI report discusses CJD, it is vCJD that has been associated with BSE.