Dicing with Death: Chance, Risk and Health

Stephen Senn, Ph.D.

Cambridge, UK: Cambridge University Press, 2003, 251 pp. $28.00

Reviewed by:

John D. Nagy, Ph.D.

Associate Professor of Biology

Scottsdale Community College

Scottsdale, AZ

Whether one loves, hates or is indifferent to statistics, it is impossible to avoid the fact that scientific medicine is founded on statistical thinking. Statistics does much more than help us determine if a drug or treatment procedure is efficacious. It provides a rational framework for studying a stunningly broad array of medical questions. Will an annual mammogram measurably decrease a 40-year-old woman's risk of early death? What is the best public health response to minimize morbidity and mortality from an intentional release of *Bacillus anthracis*? How can we rationally allocate limited medical and financial resources to deal with public health risks into the future? All of these questions, and most associated with public health and even curative health care, ultimately require quantitative evaluation. Dig down to the epistemological heart of medicine, and there you will find statistics and mathematics smiling, some might say glaring, back at you.

But, that is a good thing. A small proportion of humanity has learned a great deal about mathematics and statistics and their applications to medicine. Unfortunately, the general public is largely unaware of these efforts mainly for two reasons. The first is an epidemic of mathematical ignorance affecting even some of the most highly educated. The second reason stems from the most common type of mathematics applied to medicine, namely probability. The human brain has the unfortunate property of tending not to understand random processes intuitively but believing that it does. Don't agree? The border between the Great Basin and Mojave deserts contains arguably the harshest climate in North America. People from all over the United States, most of whom work hard and hate taxes, go to a little spot right in the middle of that hell-on-Earth and happily leave their money there. That place is called Las Vegas. Q.E.D.

This situation creates a real problem for any author attempting to write a mathematical book for an educated lay audience. Since in that audience those who do not know much or care for mathematics (call them members of "group A") far outnumber those who at least tolerate it (call them "group B"), very few truly mathematical books for the nonspecialist are ever published. One gets either a trade book entirely devoid of mathematics, like Hawking's *A brief history of time*, which is great for group A but tends to frustrate group B, or a whiz-bang academic tome like Halder's brilliant *History of mathematical statistics from 1750 to 1930*, which is anything but brief, stuns group A into a bovine stare halfway through the first chapter and even elicits a few whimpers from group B.

In *Dicing with death: Chance, Risk and Health*, Stephen Senn has grasped the bull by the horns, attempting such a mathematical book for a general audience. The result is a largely successful effort to balance group B's desire for detail without making group A imitate livestock. Senn describes his thesis as follows: "In this book we shall look at the role that medical statistics has come to play in scientific medicine. We shall do this by looking not only at current evidential challenges but also at the history of the subject." The result is a series of examples, most of which involve health topics, together with historical vignettes about the mathematicians and statisticians who originally developed the ideas. Do not misunderstand me. Senn does not shy away from developing at least some rudimentary mathematical detail. But, for those who are staunchly in group A, Senn marks the more technical sections with an asterisk. One can skip these sections and the book remains coherent; however, doing so does some violence to an understanding of Senn's main themes.

To develop a feel for the book, consider a story problem based on, but not identical to, one of Senn's earliest examples.

> Four hundred people with high systolic blood pressure begin an experimental drug treatment. After a year, the systolic pressure of 321 has decreased into the normal range, but two subjects died from non-accidental causes. Was the new drug effective? Was it safe?

At first glance, answers like "apparently so" to the first question and "probably not" to the second may seem perfectly reasonable. After all, well over half the subjects showed improvement, and if the treatment was safe, then no one should have died.

Senn does an excellent job showing why one cannot answer either question given the information at hand for reasons that are rarely obvious to the nonspecialist. First of all, it is easy to make a common but inaccurate assumption in situations like this, which I can illustrate as follows. Suppose we measure a person's systolic blood pressure, and it turns out to be anomalously high. Further suppose that our subject makes no changes in her lifestyle. After a period of time we measure her pressure again. We should not be surprised if the second measurement is slightly different from the first because we recognize that a person's blood pressure varies over time. However, whether or not we expect the second measurement to be close to the first, without further information we tend to assume that it is an even bet whether her blood pressure has increased or decreased. But, this second assumption is dead wrong. It is not that it *could be* wrong, but that it *is almost certainly* wrong. In fact, the more likely event is that her blood pressure has decreased.

As Senn explains, this fact results from a famous phenomenon called regression to the mean. If at some point in time you happen to have blood pressure considered extreme – either higher or lower than the normal range – then in the future it is more likely that your blood pressure will be closer to normal than more extreme. This statement is no guess; we *know* blood pressure regresses to the mean because if it did not medical journals would be publishing an increasing number of reports about otherwise healthy people whose blood pressure reaches astonishingly high or low values for no apparent reason.

Because the study in our story problem only focuses on hypertensives, we expect more than half of the subjects to enjoy a "spontaneous" decrease in blood pressure, caused simply by regression to the mean. Therefore, we cannot tell using this study design if the drug had any effect.

The story problem also raises another issue Senn deals with in various portions of the book. Two people taking this drug died in a year. Is that evidence that the drug is not safe? Certainly not. Senn's argument starts with economist and general intellectual powerhouse John Maynard Keynes' quip, "In the long run we are all dead." To this Senn adds, "eventually the long run arrives" and "some of us have been running quite a while." Therefore, "as the product of [the number of] persons studied and the [length of] time they are studied increases then, under any conditions whatsoever, some of them will die *almost surely*" [Senn's italics]. Since we have no information about how far into the long run our subjects are – that is, how old and healthy they are – there is no way to estimate how many of the 400 we would expect to have died even if the drug were both perfectly safe and effective.

In addition to these sorts of arguments, in the more technical sections Senn shows how one goes about calculating things like the probability of two deaths in a random sample of 400 people, or the probability that hypertension improves purely by chance (no drug effect) in 321 of 400 randomly chosen hypertensive people. These calculations are standard fare in university introductory probability and statistics courses, and anyone with at least one statistics course under his or her belt should find the mathematical development very approachable. Even without such a background, a bit of effort and applications of algebra bring the main ideas within reach.

Perhaps Senn's most innovative contribution is his attempt to acquaint a lay audience with an "argument," although that is not quite the right word, among modern probabilists and mathematical statisticians. The debate focuses on the very foundations of probability – what does "probability" mean? There are two primary schools of thought: the Bayesian and the frequentist, sometimes termed classical, interpretations. As an introduction to this very important distinction, consider the phrase, "a fair coin," which we understand to mean that, if flipped, the coin has a probability of 1/2 of landing heads-up. But, this understanding cannot serve as a *definition* of probability, at least if we insist that definitions not be tautologies. We also cannot define the probability as the proportion of times the coin lands heads-up in repeated trials; do we really expect a fair coin to land heads-up 2.5 times in five tosses? The frequentist interpretation gets around this problem using the concept of a mathematical limit. Specifically, the classical definition of probability is the exact *proportion* of times the coin will land heads-up in the limit as the number of trials goes towards infinity. On the other hand, the Bayesian interpretation allows probability to be much more fluid, and introduces two different concepts of probability – prior and posterior probabilities. Roughly speaking, the prior probability estimates your best bet given the information at hand, but after observing some finite number of trials one revises this best bet in a strictly rational way to obtain the posterior probability.

The frequentist interpretation, while extremely fruitful, has some difficulty with questions like "what is the probability of obtaining exactly two heads in four coin flips?" The hypothetical strict frequentist would suggest that the answer is 3/8 *if the coin is fair*. Senn shows how to obtain this value in a couple of different ways, using a counting technique and the binomial formula. But to complete any such calculations using the frequentist approach we have to assume something about the probability of obtaining a heads on a single coin flip. Without that information the hypothetical frequentist is dead in the water.

The Bayesian approach, however, offers a solution. A hypothetical Bayesian faced with the same question would acknowledge that further information is required and then calculate an estimate of the required data. One way to do this is to apply the Principle of Insufficient Reason, proposed by the French scientist and mathematician Pierre Simon de Laplace who you will learn more about in Senn's book, probably even if you already know something about him. Laplace's principle suggests that our best bet, in the absence of information about the coin, is to assume that the probability of a heads is equally likely to be anything between 0 and 1. Then we flip the coin a few times, plug the results into a formula derived by Laplace and obtain a revised estimate of the probability of a heads. For example, suppose you have a coin that you know can land either tails- or heads-up. If you then flip the coin and obtain a heads, then your best bet for the actual probability of a heads is 2/3 according to Laplace's procedure. No one claims that the probability of a heads *is* 2/3, just that 2/3 is our best estimate under the Principle of Insufficient Reason given the results of a single coin toss. Of course, more data (coin tosses) allows Laplace, in the guise of his formula, to produce a more refined estimate.

In this example one may balk at the assumption called the Principle of Insufficient Reason. Just given their physical nature, we would usually expect an unfair coin to be nearly fair, yet Laplace's principle insists that we consider "heads is wildly more probable than tails" to be just as likely as "heads is just slightly more probable than tails." But, as Senn points out, "in general, there is insufficient reason for the principle of insufficient reason." This fact need not cause involuntary muscle spasms because the Bayesian approach does not rely on the principle; it is only used for historical reasons and to allow us to perform the calculations in the example.

At this point many readers of this journal may be wondering what all this lofty (lowly?) discourse has to do with medical ethics. The general answer is, "a lot." More specifically Senn's discussion of this "debate," and the Bayesian view in particular, is important because most introductory university statistics courses, beyond which the majority of students never get, are taught almost strictly from the frequentist's perspective. In such courses one learns how to calculate the probability of experimental or survey data given the probability structure of the problem. By "probability structure" I mean parameters. But the Bayesian approach looks at this backwards – one attempts to determine the parameter(s) given data. Anyone interested in epistemology of scientific medicine really needs to know about the Bayesian school but will not have an easy time finding approachable sources describing it, at least until now.

As an example, although one not used by Senn, suppose a particular patient "Joe" presents with a collection of signs and symptoms, some of which are consistent with diseases A and B, others of which indicate A and not B, and still others suggest B and not A. A test for disease A is ordered and comes back positive, but we know that this test gives false positives 2% of the time. What is the probability that Joe has disease A and not B? From the frequentist perspective no such calculation is possible because the probability is undefined – Joe either has A, or he does not. However, the Bayesian approach allows one to calculate a (perhaps) meaningful probability for Joe suffering from disease A.

In addition, the Bayesian approach offers the ethicist fuel for some very interesting discussion about clinical trials. The double-blind, placebo-controlled, randomized clinical trial is widely viewed as the gold standard test of efficacy. But as Senn points out there are very serious ethical issues associated with assigning individuals, as we must in such a trial, to the placebo group, even if they still receive the standard of care. The common interests of researchers and future patients are best served if some current patients take a placebo; however, current patients will tend to perceive that taking a placebo is not in their best interests. Senn motivates this dilemma in detail, and with an example and an appeal to his references, he indicates situations in which one might use a Bayesian approach to design a study that somehow balances the interests of current and future patients in the "best" way.

There is much more to this book than the few examples provided here. Senn treats a variety of very important and interesting topics, including the devastating inefficiency of clinical trials that include both sexes and all races. That's not to argue that one should only study one sex and race, but rather that in a *single trial* the more variation one adds, the less likely that trial is to uncover a pattern for a given number of subjects. Senn also includes chapters on life tables, meta-analysis, statistics in law and the current debate surrounding the MMR vaccine and autism. He even includes a short treatment of mathematical models of epidemics, a topic foreign to books on medical statistics but not at all unwelcome. Throughout the book Senn mixes history, practical applications and some technical detail in a thoroughly engaging prose. In fact, perhaps the best thing going for this book is the charming wit and style of its author. (Anyone who associates Switzerland with cuckoo clocks, beware.)

Unfortunately, nothing is perfect. This text suffers from an almost complete lack of editing beyond the author's. In fact, there's little evidence of either style or copy editing, even to the extent that one might expect from reviewers. There are a number of typographical errors, including both text and mathematical notation, some of which can lead to confusion among those who are near their mathematical limits. And, although the historical portions add enormously to the book, their integration needs rethinking. In a number of cases, Senn begins developing an idea, then cuts to a historical section largely irrelevant to the direct development of the concept, and then returns after we have forgotten what the original topic was. Such a defect can be easily corrected with a minor reorganization one would hope an editor would suggest. As it stands, anyone starting to read the book should realize that they have the author's manuscript in hand, not a thoroughly finished product.

Normally, a level of tolerance for such failings in an academic book from an academic publisher is the rule, but in this case I am afraid it may turn potential readers away. Since a significant fraction of the book's intended audience may be harboring some trepidation about the subject matter before they even begin reading it, I fear that the rough spots will simply cause unnecessary frustration that will discourage readers from finishing. And that would be a real shame because not only is the work very charming and enlightening, but if I may borrow Senn's own words, "[i]f you think that statistics has nothing to say about what you do or how you could do it better, then you are either wrong or in need of a more interesting job."

References

Anders Hald, *A History of Mathematical Statistics from 1750 to 1930*, Wiley, 1998.

Stephen W. Hawking, *A Brief History of Time*, Bantam Books, 1988.