Why I don't read medical literature

Mark Elliott, MD, FRCPC

Medical evidence has a credibility problem that is rooted in the fundamental problems with statistics. This problem is manifest in the inability to reproduce evidence on repeated randomized clinical trials (RCTs). Theoretically, an RCT is the way to answer questions about which treatments are useful, but practically, an RCT is too expensive to conduct with enough patients to get answers to such questions.

There are three fundamental problems with statistics. First, in 2005 John Ioannidis wrote what is now very well cited paper in *PLOS Med* entitled, "Why most published research findings are false."[1] In the article he likened medical research to a large RCT machine with three dials that can be conservatively set to show that medical research is 95% wrong. The first dial shows how many more false hypotheses there are than true ones. The second dial shows how underpowered studies are. The third dial shows how much published results are flooded with false positives.[2]*

Second, there are very precise impressive algebraic statements made about the bell-shaped curve.† These statements have names like *variance* and *confidence intervals*. These statements simply tell us how uncertain we are about the average treatment value we measure from the data generated from our experiments compared to the real treatment value we would have if we had all the data, which we rarely do. Furthermore, biology rarely con-

Dr Elliott is a staff anesthesiologist at Providence Healthcare in Vancouver.

forms to a bell-shaped curve.[3]‡ This means that all those impressive algebraic statements about how uncertain we are about our treatment are completely and absolutely meaningless.

Third, probability theory was originally developed for nothing more than winning at games of chance. Origina-

> **The statistical language of science is very different from the cause-and-effect language that the human brain uses.**

tors of this theory§ only cared if cards were dealt fairly and unpredictably to the players. There was no assumption of any random process going on. This leads to normative statements. Normative statements are very strange for a scientific enterprise. You can say that a full house beats two of a kind. That's true, because this is an "is" statement. However, the statement that you should not draw to an inside straight is a normative statement, which is an "ought" statement. And you cannot get an ought from an is.‖

Luckily there are three improvements taking place. First, you can now preregister a study with a reputable journal that does not charge a fee for publication (some regrettably

do). If accepted, and as long as you follow your protocol, the study will be published whether the results are positive, negative, or indifferent.

Second, people are discussing cognitive biases more openly. There is now a large body of published work in the fields of evolutionary psychology and behavioral economics that basically says the world does not run on rationality. Hundreds of these biases are now catalogued. However, the obvious which-drug-company-is-paying-for-dinner cognitive bias usually isn't discussed. This means that which treatments get promoted are still, to a large extent, based on marketing.

Third, there is now a causal revolution underway in statistics.[4] The statistical language of science is very different from the cause-and-effect language that the human brain uses. As we came out of the trees and onto the plains of Africa, a part of our brains that had evolved to detect patterns in nature was co-opted to detect cause and effect. These early monkeys with big brains realized that A was followed by B and that one could manipulate A to change B. The mathematicians who developed statistics in the early 1900s did away with cause and effect so as to deal only with correlations because it made the math easier. There are now well-established causal inference tools that will help researchers make predictive statements from descriptive data alone.

Hopefully this credibility problem will improve. Hopefully our academic institutions will be more upfront

in their push to have students do research as a means to a job. However there is a very large tyranny of the status quo to overcome.

References

1. Ioannidis JPA. Why most published research findings are false. PLOS Med 2005;2:e124.
2. Muller D. Is most published research wrong? YouTube. Accessed 4 April 2019. www.youtube.com/watch?v=42QuX LucH3Q&t=266s.
3. Wolchover N. At the far ends of a new universal law. Quantamagazine. Accessed 10 April 2019. www.quantamag azine.org/beyond-the-bell-curve-a-new -universal-law-20141015.
4. Judea P, Mackenzie D. The book of why. Basic Books; 2018.

Footnotes

* Derek Muller went to West Vancouver Secondary School and wanted to be a film director, but wound up at Queen's University, Kingston, for a BSc in engineering physics, then the University of Sydney for a PhD in physics education research. He has over 5 million subscribers to his YouTube channel, Veritasium, which produces science videos.

† A bell-shaped curve is also known as a normal distribution or a Gaussian distribution. If you go out two standard deviations from either side of the mean you account for 95% of the events, leaving only 2.5% in each of the tails of this type of statistical distribution. This is well described with mathematics, but the mathematics is only an approximation. It is a good approximation but not perfect. Physicians are usually quite in awe of all this math and tend never to question the underlying assumptions. But math is just another language, like English or French or any other language. The language of math is pretty good for physics but terrible for biology.

‡ There is a statistical distribution made famous by Nassim Nicholas Taleb's book *The Black Swan*. It is a distribution that is just a little lower in the centre and just a little higher in the tails. So all the action is now in the tails. If one overlays a fat-tailed distribution on top of a Gaussian distribution one can barely make out the difference by eye. But the probability of an event 10 standard deviations from the mean (a 10 sigma event) is one in a trillion in the Gaussian (remember the 2008 financial crisis apologists) but one in a hundred in the fat-tail distribution. This means there is a nine orders of magnitude change in those precise algebraic statements that tell us how uncertain we are about our measurements. Human brains have difficulty understanding such exponential changes. For example, a six order of magnitude change would be buying a $2 million house in Vancouver for $2. Distributions for systems in which many variables are interacting (like in biology) probably follow an asymmetrical distribution called the Tracy-Widom distribution. If this is the distribution underlying our science, then those statements of uncertainty are not just orders of magnitude wrong, but simply meaningless.

§ One of the pioneers in this field was a physician named Gerolamo Cardano (1501–1576) whose book *Liber de Ludo Aleae (The Book on Games of Chance)* includes a chapter on cheating.

‖ This ought-from-an-is debate has been going on for hundreds of years. At the start of the scientific revolution, David Hume wrote about it extensively. Yet despite this underlying philosophical paradox in probability theory, probability has taken over vast areas of human endeavor such as the actuarial science used by the insurance industry and the Black Scholes model of pricing financial derivatives (and we know how that turned out in 2008).