

Alberta Doctors' Digest

Model mistrust and Bayesian blues

“Medicine is a science of uncertainty and an art of probability.”

- Sir William Osler

I've been thinking about models these last few months – the statistical ones, I mean, their complexity – with skepticism and growing mistrust when modellers are cagey about the murkiness of their projections, especially when these lead to huge impacts on people's activities and a nation's economy as we are swept into the biggest wealth transfer from taxpayers' savings to government debt since the two to three decades following World War II.

When I was nine, I liked models. I bought a kit and built a paper model of the Houses of Parliament, perfect in scale, about the size of a table mat. I knew the real Parliament was much larger and built of stone, not paper. Twenty years later I noticed there were details that had not been accurately depicted, but the difference between my model and reality did not disturb me. I understood what was going on.

The current problem is not about attempts at statistical modelling – although the complexity engenders some mistrust through methodology suspicion – but is more about some modellers' tendency to bend, twist and stretch their interpretation based on their own biases. In this era of media disinformation and fabrication spread by politically motivated or credulous activists, it's important for one to understand and trust information.

In Nanaimo in mid-March, I was playing golf. My old friend Alan and I were physical distancing and not touching the bunker rakes. On March 17, [Imperial College Report 9](#) was released and Professor Neil Ferguson's predictions hit the headlines. That contributed to Nanaimo Golf Club closing and us high-tailing back to Calgary for access to a ventilator. Professor Ferguson (now known in the U.K. as “Professor Lockdown”) is a mathematician and epidemiologist, not an infectious disease expert. His *Report 9* (with 30 fellow authors) predicted there could be over 510,000 deaths in the U.K. between April 20 and August 20, over 2.2 million in the U.S.A. and over 40 million worldwide. That was what hit the world's media. *Report 9* makes interesting reading in retrospect although few actually read this (non-peer-reviewed) report, and the media took hold of these stunning numbers:



In western Canada, the catastrophe did not happen but led to many urgent procedures and therapies being delayed (photo credit: Pete Linforth, Pixabay.com)

“We find optimal mitigation policies (home isolation of suspect cases, quarantine of those living in the same household, social distancing of the elderly and others at most risk of severe disease) might reduce peak health care demand by two-thirds and deaths by half. However, the resulting mitigated epidemic would still likely result in hundreds of thousands of deaths and health systems (most notably intensive care units) being overwhelmed many times over. For countries able to achieve it, this leaves suppression as the preferred policy option.”

Ferguson should have forcefully emphasized that this model was an extreme scenario of doing nothing and assuming an exponential spread of a uniformly lethal infection with an R (number of other people one case is likely to infect) of 3.0. In the U.K., politicians were horrified. It escalated the U.K. lockdown with knock-on effects in the U.S.A. and other countries. Another consequence of Ferguson’s visit to 10 Downing Street was likely the transmission of COVID-19 to British Prime Minister Boris Johnson.

Several groups offered different predictions to this apocalyptic scenario, using lower case-fatality rates and lower R rates. These included the Oxford Evolutionary Ecology of Infectious Disease Group who suggested Ferguson’s prediction was an extreme of possibility and not a probability. They showed a vastly different outcome using an R of 2.2 rather than 3.0 (as in the Imperial forecast).

It’s not the first time Ferguson has been criticized. Back in the early 2000s, he predicted 136,000 human deaths in the U.K. from mad-cow disease (there were maybe a few hundred) which led to six million cattle being destroyed, which decimated the rural economy. This was brought up by Edinburgh University’s Professor Michael Thrusfield, an animal epidemiologist, who talked of “d  j   vu” recalling Ferguson’s over-prediction of human deaths following bovine foot and mouth disease and urged caution on this latest prediction. Then there was his over-prediction of the effects of bird flu. *Report 9* also predicts a further six waves of COVID-19 through to August 2021. Ferguson could be right for a second wave given the anti-racist demonstrations in early June in so many cities, but further waves will likely involve fewer cases since physical distancing, face masking and hand cleansing will be rapidly re-instituted. Other groups have pointed out that the software from Imperial’s individual data model is over 13 years old. So Ferguson’s star is fading – particularly after being caught evading the lockdown he himself proposed by visiting a girlfriend several times – perhaps he was physical distancing. There’s an old Polish saying: “You never know unless ...”

Models in Canada predicted lower death levels. In Alberta, even these lower predictions meant a massive reorganization of patients and beds to accommodate the potential crowds of sick and dying. In western Canada, this catastrophe did not happen but led to many urgent procedures and therapies being delayed. On the other hand, it had the effect of galvanizing reluctant governments into lockdown action. But crying wolf has a diminishing efficacy.

$$P(A_i|B) = (P(B|A_i) \times P(A_i)) / (\sum_{i=1}^n (P(B|A_i) \times P(A_i)))$$

Basic equation of Bayes's Theorem (this would be a good name for Elon Musk's second child, if he has one)

The statistical predictions used by epidemiologists often use a "Bayesian" approach.

Bayesian? Named after the Reverend Thomas Bayes (1703-68), an English Presbyterian minister who bettered himself by studying logic and divinity at Edinburgh University. Bayes was intrigued by philosopher David Hume's proposition that miracles should not be believed solely on anecdotal evidence. Bayes became interested (even obsessed) with the problem of the probability of an event occurring and devised a mathematical formula to measure that probability. Gamblers have taken an interest in this branch of statistics ever since.

Until the 1950s, statistics was dominated by frequenters (statisticians manipulating numbers of observed events). Most of the statistics used in medical research (e.g., clinical trials results reporting) are frequentist.

All physicians practice the art of probability assessment, constantly assessing the probability of a patient's diagnosis, the response to therapy, the possibility of having side-effects, dying or getting better – we're all probability mavens at heart, but statistical Bayesians we are not.

Bayesian statistics are usually explained by the example of flipping a coin and predicting the likelihood of heads or tails, modifying the probability of getting a 50:50 chance by number of tosses, the strength of the toss, whether the coin is bent, etc. If an event (A) is influenced by a factor (B), there is a variable influence. This influence depends on "conditional probability" formulae where the probability of the event occurring is influenced by the probability of other events or influences occurring which may influence the final likelihood of the event actually occurring. And we won't dig deeper into Gibbs sampling or Markov Chain Monte Carlo modifications because then my head starts spinning. I'm no statistician, but I can see the fragility of all this probability juggling.

To build a model to predict events in a population suffering an epidemic, you have to give a numerical weight to a number of demographic factors such as age, sex, race, social class, income, population density, general health, weight and medical complications, then update these with factors that become evident through observation such as proximity to an international airport, number of night clubs, etc. These then are given a relative value and matched to other numerical assessments of, say, virus

infectivity, susceptibility or disease severity. All Bayesian statistical calculations require constant updating as information comes in on real R rates, ICU admissions, ages of patients dying and population behaviors.

The complexity is mind-blowing and falls prey to incorrect relative valuations. King James I might have been talking about Bayesian statistics, and not John Donne's poetry, when he said: "Dr. Donne's verses are like the Peace of God, for they passeth all understanding."

In 2014, I participated in a group led by an American statistician adept at Bayesian statistics whose company had been commissioned by a drug company to predict what increment of HER-2-positive breast cancer patients achieving a complete pathological cancer response (pCR) with HER-2 directed neo-adjuvant chemotherapy (chemotherapy given before surgery) would be required to show a statistically significant survival advantage (and thence gain FDA and HPB approval and mouth-watering profits). We used previously published trials data.

Our model predicted that around a 10% increase in pCR rates should be enough to predict enough of an increase in survival to convince the FDA and HPB to approve a new drug. [I was an author on that paper](#). I could not follow the statistical manipulations but went along with the paper because my gut feeling was that the conclusion was correct and besides, I'd already cashed the cheque. That's weak, but the traditional schoolboy last gasp excuse (which I have used frequently in this life) is: "I wasnae the only one, sir."

I have support for this cowardly decision from Sharon McGrayne in her book [*The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*](#): Laplace built his probability theory on intuition: "...essentially, the theory of probability is nothing but good common sense reduced to mathematics. It provides an exact appreciation of what sound minds feel with a kind of instinct, frequently without being able to account for it."

With the COVID-19 pandemic, everyone's a Bayesian now – most citizens understand how we must "flatten the curve" and reduce the R rate to reduce pressure on ICU beds by physical distancing and hand cleansing. Vanishingly few understand how these curves are produced and with what flimsy scaffolding they are constructed – by necessity – since for this latest pandemic, initially at least, its natural history was unknown.

This amiable statistician said at one point in our meeting: "The Bayesian statistical approach is difficult in the sense that thinking is difficult."

The others smiled or laughed knowingly. I took comfort in my colleagues thinking they understood more than they probably did. As one striking modeller, Elizabeth Hurley, model, actress and business woman, has said: "I would seriously question whether anybody is foolish enough to really say what they mean. Sometimes I think civilization would break down if we all were completely honest."

One tiny part of our paper's statistics section read:

“For analyses, the joint posterior distributions of model parameters are fit using Markov Chain Monte Carlo using the estimated number of events and exposure time by pCR group within each time segment. We sample values of model parameters from the joint posterior distributions using Gibbs sampling. When closed forms to full conditionals are unavailable Metropolis-Hastings is used. When studies provided only the total number of events across all time segments, we impute the segment-specific number of events after each iteration of the MCMC from a multinomial distribution using the current sampled segment-specific baseline hazard rates, λ_t , and the amount of exposure within each segment.”

I asked: “Is this the kind of modelling for predicting climate change?”

There was laughter from the statistics folk. “Yes, of course.”

The problem is that the fallibility of Bayesian modelling is not emphasized. Accuracy depends on accurate relative numerical weighting – you don’t really know what goes into it – given the host of unknown unknowns or even known unknowns. It’s like the roadside food-trucker who was asked how he could sell rabbit sandwiches so cheap. “Well,” he said, “I put some horse meat in too. But I mix them 50:50. One horse, one rabbit.”

A better word than “probability” might be “possibility.” Next time there will be more skepticism. But what we need is not more skepticism – there’s enough of that – but more understanding of the weaknesses of Bayesian epidemiological statistics, with the promise of rapid improvements as new data comes in.

Will the population be as credulous the next time, knowing the massive debt and the limited risk in less populated areas? I doubt it. Some trust in modelling has been lost. Some mathematical modellers fall prey to the attraction of celebrity and press attention. All models are fallacious, but all can be useful with caution.

What has not been lost is trust in the wisdom of distancing, hand cleansing and perhaps face masking.

Editor’s note: The views, perspectives and opinions in this article are solely the author’s and do not necessarily represent those of the AMA.

Banner image credit: Chuk Yong, Pixabay.com