

Prediction impossible! Québec's mortality predictions a week later confirm politicians and public servants need data science training

14 April 2020

By Anessa L. Kimball,

Professeure agrégée de science politique & directrice, Centre sur la sécurité internationale, ESEI, Université Laval, also Co-directrice, sécurité, Réseau canadien sur la défense et la sécurité (CRSH) & member, North-American Team, Defence Foresight Group

In the midst of crisis, information transmission and clarity remain essential as governments deliver policy goods and manage the externalities of the accumulating pressures upon the public apparatus. The current crisis is a collective action issue involving public health, specifically, its success or failure. The failure of collective action results in a public health challenge larger & more complicated than most societies have faced. It involves levels of governance and collaboration across stake-holders with differing levels of power and access to information. As a natural experiment, "covid19 pandemic" presents many challenges economists and scholars of collective action theory have explored for decades, since the 1960s work by Olson¹ and 2000s extension by Ostrom on the Tragedy of the Commons, also Nobel Prize winning in 2009², illuminated multiple challenges associated with preventing collective failures. In 2004, Sandler's book *Global Collective Action* offered an entire chapter on the challenges of global and transnational health issues using the illustration of pandemics. As a researcher and professor with two decades of experience examining political institutions, collaboration, economic analysis, and defense/security using statistical modeling, I have a set of expertises enhanced by personal experience, as a cancer survivor since 2015. I underwent 2.5 years of isolation; I am a datapoint in multiple (cancer) survival models with the obligatory research consent paperwork.

If Legault and PM Trudeau had mobilised collective action arguments and ideas, the current informational state in Canada and Quebec would be different. If politicians and civil servants understood data science and multi-level simultaneous modeling from the beginning of the crisis, the public discourse would have been nuanced. As an experienced data scientist, the signals sent by several legitimate speakers in Canada, represent a level of "catching-up" concerning statistics which is worrisome as both a resident and a public intellectual. Citizens, concerned with the terminal state of the illness, morbidity, placed pressure on politicians to offer estimates. Despite politicians possessing better information, they were still placed in a "mission impossible" scenario. One week ago, the QC government offered is "predictions".³ A week of retrospective provides some distance to evaluate those predictions based on data science.

The increased/better information politicians and experts have include data: on federal and interprovincial transmission rates (as well as individual spread vectors), more information concerning 1) differential health care capacities/potentials across provinces (semi-private info about beds, Drs, nurses including those working part-time/recently retired), 2) if citizens are truly isolating (from social media/phone data the government can request from companies), 3) estimates of those with symptoms but not tested because they are in low risk categories who were told to self-isolate unless medical help required as well as indirect information (via internet searches for covid19 symptoms/pharmacy purchases for medications). In addition, politicians and experts understand the transmission of the virus is affected by urban geography and population density making national level predictions more problematic. Though asymmetric information is a major issue, I contend a more important issue is the **lack of basic data science knowledge among politicians and civil servants** which results in impossible predictions being offered as "genuine/legitimate" knowledge to the public.

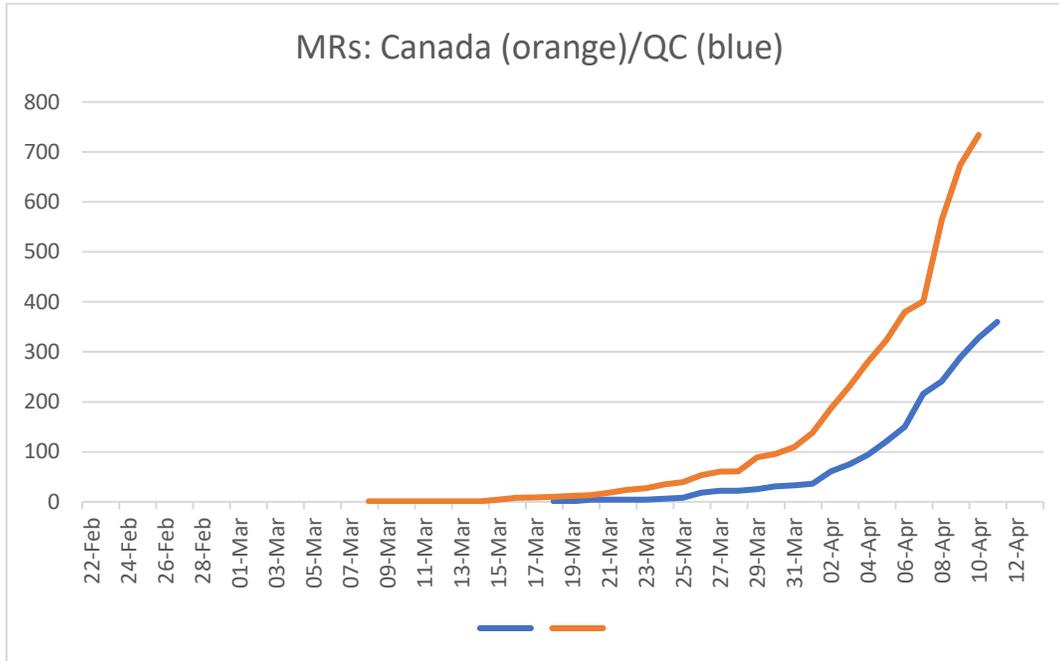
Data science includes not only the raw data (morbidity rate) but also the ability to implement the appropriate statistical model including the factors related to changes in morbidity. First, understanding a morbidity model properly is a sequential process, involving three phases, as Canada is only testing symptomatic individuals. If one is symptomatic,

¹ *The Logic of Collective Action*, 1965. Olson was nominated for a Nobel Prize for a fundamental contribution to social bargaining for the aforementioned work.

² <https://www.nobelprize.org/prizes/economic-sciences/2009/ostrom/facts/>

³ <https://www.lapresse.ca/covid-19/202004/07/01-5268428-de-1263-a-8860-morts-dici-le-30-avril-selon-les-scenarios-de-quebec.php> and Scénarios: évolution de la COVID19, 7 avril 2020, Québec.

does one get tested → if tested, is one positive for the virus → if positive, does one die. A model such as this is hierarchical, so built at the provincial level and aggregated federally. There is little evidence in the current discourse of such an approach, provincial modeling and federal modeling appears to be separate.⁴ Below is a figure comparing Canada and QC morbidity rates, note the difference in first death date. It is not clear how the federal government is integrating multiple provincial models. More problematic is the lack of transparency concerning model specification, in other words there is no indication of exactly what causal factors are in the models, again noting the above paragraph offers a slice.



Second, the federal modeling provides no special social indications for Quebec. At no point has the typical greeting of cheek-to-cheek kisses been identified as a vector for the social spread in Montreal and eastern areas to explain the overspread in Quebec and the transmission at the ON/QC border areas. On Good Friday, Trudeau offer a public acknowledgement of areas of high and low contamination indicating his understanding is improving of the nature of the pandemic. The other slices of causal factors include socio-cultural, demographic, local, police actions, etc. The objective herein is to present the “absence of science” in the QC mortality rate prediction and demonstrate “expertise” is in the eye-of-the-powerful⁵.

As a data scientist and collective action bargaining (failure) scholar teaching a graduate course on international institutions which includes studying the World Health Organisation. As early as February,⁶ we began discussing the issues setting up the world for collective failure of managing the pandemic. Early arguments included the structural and financial weakness of the WHO as an institution, including instable budgeting, the lack of a central secretariat, and a de-prioritization by states and society. When I joked to students, it was the delivery team for Bill Gate’s malaria nets, it drew some nervous laughter, though none of 25 students named the head doctor for the W.H.O. in a growing pandemic. As the term continued, we discussed what second wave spread could resemble. The average age of the cases would decrease due to social beliefs it was a “boomer illness” and “youth immunity.” Also, how social inequalities could increase the cleavage. Then Italy, got sick. When Italy got sick, discussions changed since so many people have connections to Europe and several students in my class were from there. Sandler’s book being required for the course, we discussed in-depth how culture, rules, institutions, and social actors both within and across societies, containing differing enforcement capacities, accesses to information, and rules of law, must collaborate to pull off a “collective action” win. In other words, 25 individuals (educated from all over the world with a diversity of undergraduate degrees)

⁴ Federal modeling document shows cases all starting at the same date, though empirically untrue.

⁵ A large literature on the social constructivism is available. Wendt’s 1992 *International Organization* research article “Anarchy is what state’s make of it,” provides a departure point.

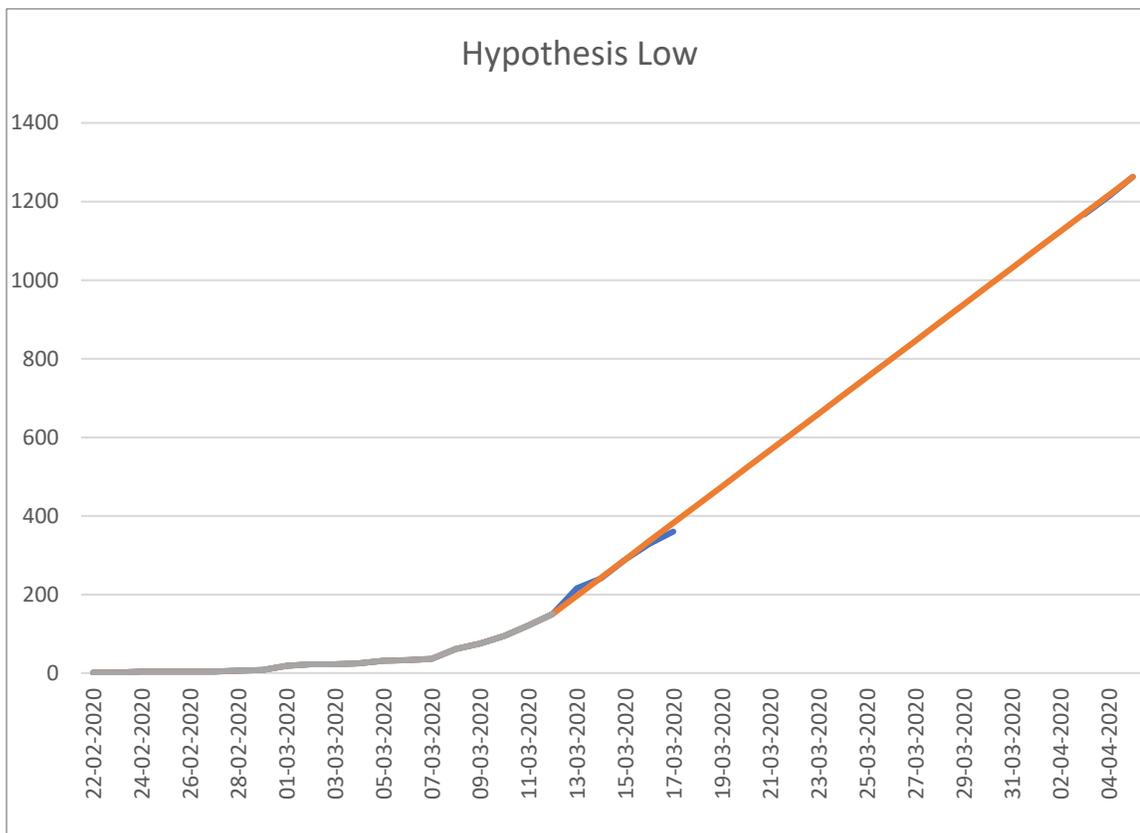
⁶ <https://twitter.com/caixin/status/1225681410033254400?s=20>. I retweeted this @ProfKimball on 7 February 2020.

discussed this exact challenge⁷, all of the various factors to put in a predictive model of (international/national collective action success/failure) and we all felt, our model needed more work.

On the 7th of April, the QC government released predicted morbidity rate range possibility for the 30th of April. “De 1263 à 8860 morts d’ici le 30 avril, selon les scénarios de Québec. »⁸ and as a data scientist following (for months) the observed/reported numbers privately on the Johns Hopkins University COVID19 Dashboard⁹. Upon the “reporting” of the numbers, something more sophisticated than a “gut instinct” was called upon, it was data science. After the work shift followed by the mom shift ended, I calculated into the evening because those “predictions” served as hypotheses (estimates), as a data scientist I understood from the existing data available and was able to finish the function mentally before mathematically. The “expert” understanding about statistics in my possession, I was able to reverse the math on the low prediction (1263) before the end of the evening.¹⁰

Despite an intellectual win, the loss is the public that put faith and credibility into the government’s prediction. The QC government was understandably reticent to offer any prediction, given the better understanding of the experts behind the scenes. In other words, the impossibility to be able to put together all the factors one would require defaulted the government into offering a “simplistic/rudimentary” prediction, I discovered to my non-surprise one could predict the lower estimate as a function of nothing but passing days multiplied the average (post-exponential split, approx.) plus the QC morality rate that day¹¹. Additionally, a linear prediction is not the appropriate functional form of the relationship, i.e. it is an exponential relationship.

This is the equation for a line ($Y = aX + b$) or Expected QC mortality rate¹² = (time*46.355) plus 150.



⁷ Across 7 course meetings over 5 weeks.

⁸ See footnote 1

⁹ <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>

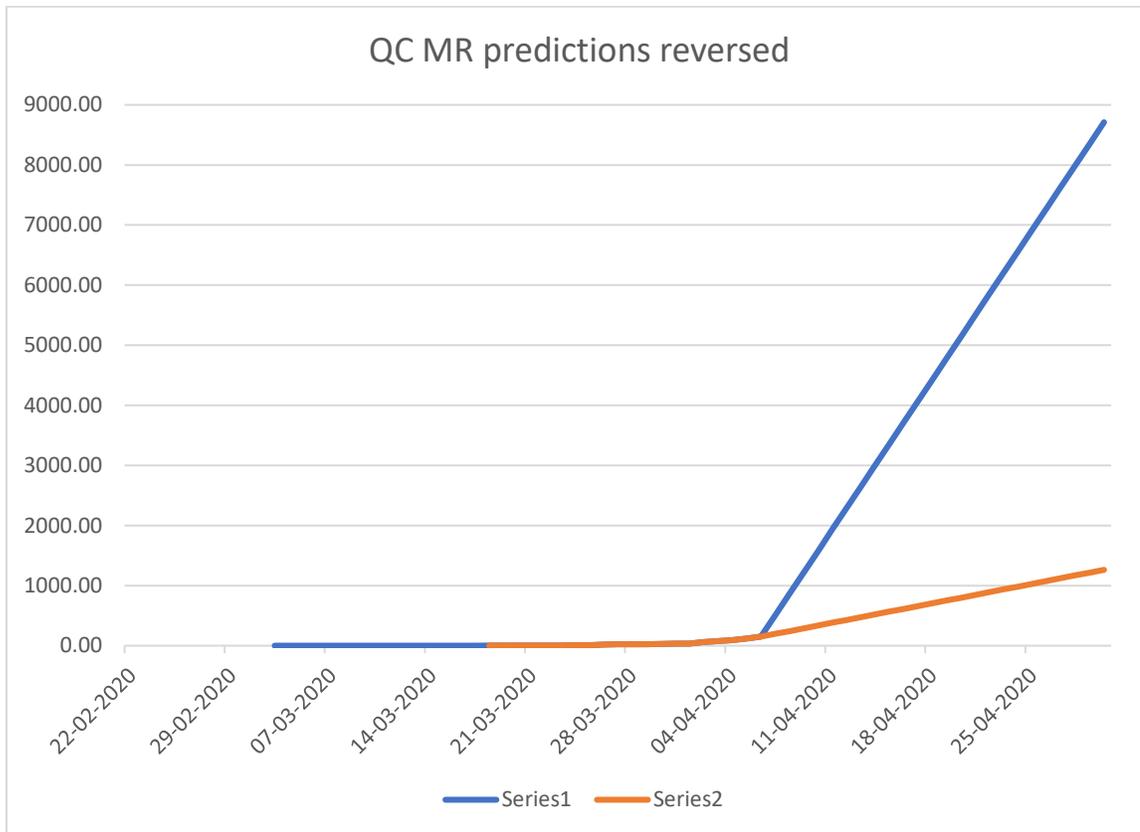
¹⁰ <https://twitter.com/ProfKimball/status/1247695392843710472?s=20> Data available upon request.

¹¹ Grey line

¹² Orange line

A realization all the more unsurprising knowing how little data science training is required in Canada in political science departments¹³. Again, the sequencing of the process and delay to morbidity suggests the last week's data (the blue line above starting at 150), possibly indicates a world without maximal social distancing interventions. If social behavior is slowing the spread to and preventing mortality, then should start to flatten out in the 7-10s or virus cycle.¹⁴

My third shift the next evening was occupied by reversing the higher prediction 8860, which I reversed again by fitting a scenario equal to the logic of "7.5-8 times worse". Expected QC mortality rate High prediction (orange line in figure below) and mathematically equals = (time*(46.355*7.699)) plus 150. Therefore, the high prediction is technically, 7.699 multiplied by the earlier values. What is not surprising, is that I reversed the QC predictions; ***what is surprising is the "politicians" had so little knowledge, it resulting in an incapacity to detect an "empty model" from a "valid causal model". That is the point we should be discussing.*** Politicians and experts have failed to present the information credibly and clearly to the public, which is a story not to be missed.



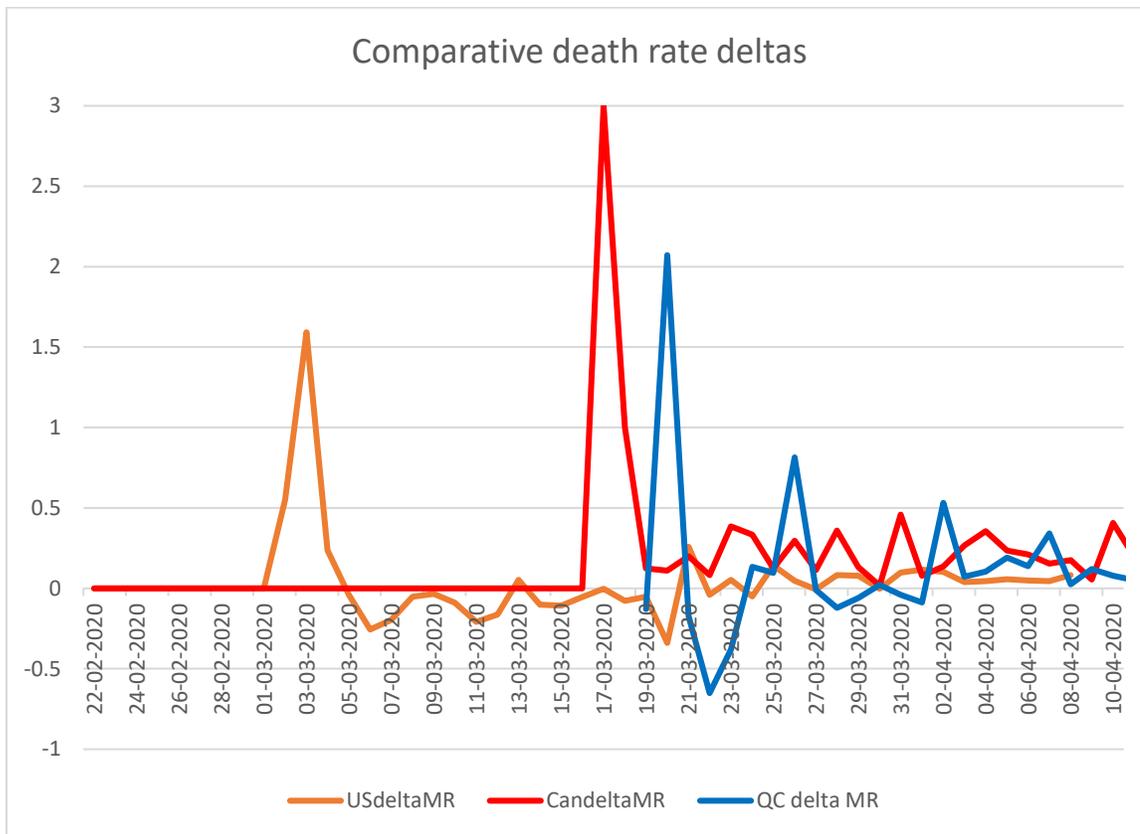
As a data scientist, economist, professor and public intellectual, to be relevant any public policy inquiry into the data and morbidity rates should offer a relevant alternative if the evidence presented confirms a lack of science. And, in particular, survival models of data analysis¹⁵ mobilize a distinct vocabulary in discussing the effects of interventions (human/medical) on the hazard of death (i.e. the rate of death). Additionally, in (cancer) survival models, there is an important attention to the differential effects of such interventions across patients due to pre-existing conditions medically (or in cancer genetically also matters to mortality rates). Understanding the metaphors and language of statistical survival models complemented by decades of studying and publishing on collective action, power, and political institutions as well as social cultures and how data is collected across countries, I offer a different measure to track at

¹³ See Dion, M. & L. Stephenson. 2017. "Planning for the future: Methodology training in Canadian universities," Canadian Journal of Political Science, 50(1): 281-294.

¹⁴ Note 10 days between first Canadian death & first QC death.

¹⁵ Despite completing the required 4 course methods sequence and the recommended 2 course summer statistics camp (ICPSR) in graduate school, I audited a survival model course to contribute to the pre-submission draft comments on Box-Steffensmeier & Jones 2004 Cambridge book, *Event History Modeling: A Guide for Social Scientists Paperback*.

multiple levels.¹⁶ The rate of change in morbidity rates offers an eloquent generalised mathematical solution drawn from local rates permitting comparisons between US (orange), Canada (red) & Québec (blue).



The simplicity of this measure is the capacity to draw an inference from the location of the point and its tendency.¹⁷ This permits identification of the deaths flattening by proxy.

Best signs: rate is consistently negative & increasingly negative

Good signs: rate is negative

Neutral signs: rate is at zero (no growth)

Bad signs: rate is non-zero & positive

Worst signs: rate is consistently positive & growing

No one can predict the total death rate but one can collect evidence and show what the tendencies are for policy making. Current world: Bad signs though QC is moving closer to zero showing signs of neutral though local level will show Montreal to be different. I am far not alone in skepticism about the predictions but the difference *is I decoded the QC expectations to demonstrate their impossibility/unscientific basis deploying the "legitimate/expert 'knowledge language' of statistics"*.

Recommendations: the models¹⁸ for the various levels of morbidity predictions & data must be available, the discourse needs to shift to a regional/local strategy for containment and collective action, all humans must accept the short-term costs/inconveniences for medium to long term (limited) restrictions to be relaxed but vigilance is required for a year or longer until the public good (the vaccine) is made available¹⁹. As Canadian forces in the US says, we are all in this "together apart" and "togetherness" remains a sub-narrative of the multiculturalism Canadians value as much as their own personal safety & national security.

¹⁶ It can also be adjusted for population density.

¹⁷ <https://twitter.com/ProfKimball/status/1247933196139978752?s=20>

¹⁸ Programming code and datasets could alternatively be made public for data scientists.

¹⁹ Sandler also offers some predictions for the conditions making it more likely (2004).