Prediction impossible! Québec's COVID-19 mortality predictions a week later confirm politicians and public servants need data science training

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

Abstract: During crisis, information transparency and transmission remain essential knowledge goods as governments deliver policy managing the externalities of the accumulating pressures on the public apparatus. COVID-19 is a collective action crisis concerning public health, specifically, success or failure of the aggregation of individual acts. The failure of collective action engenders a public broader and more complicated health challenge involving levels of governance and collaboration across stakeholders with differing of power and information. As a natural experiment, the COVID19 pandemic challenges economists and scholars of collective action, while offering observed data to evaluate success. As a data scientist, the signals sent by legitimate speakers in Canada, represent a level of catching-up concerning statistics which is worrisome. Despite politicians possessing better information, they were placed in a mission impossible scenario. A week ago, the QC government offered the requested predictions and today those predictions are evaluated using data science.

Prediction impossible! Québec's COVID-19 mortality predictions a week later confirm politicians and public servants need data science training

In crisis, information transparency and transmission remain essential knowledge goods as governments deliver policy goods managing the externalities of the accumulating pressures upon the public apparatus. This crisis is a collective action issue concerning public health, specifically, success or failure of the aggregation of individual acts. The failure of collective action results in a public health challenge larger & more complicated involving levels of governance and collaboration across stake-holders with differing levels of power and access to information. As a natural experiment, "covid19 pandemic" presents challenges economists and scholars of collective action theory explored for decades, since the 1960s work by Olson[1] and 2000s extension by Ostrom on the Tragedy of the Commons, also Nobel Prize winning in 2009[2], 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, 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 crisis' beginning, the public discourse would have been nuanced. As a data scientist, the signals sent by several legitimate speakers in Canada, represent a level of "catching-up" concerning statistics which is worrisome. Citizens, concerned with the terminal state of the illness, morbidity, placed pressure on politicians to offer estimates. Despite politicians possessing better information, they were placed in a "mission impossible" scenario. One week ago, the QC government offered its "predictions".[3] A week of retrospective provides distance to evaluate those predictions based on data science.

The increased/better information available to politicians and experts includes 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 of three phases, as Canada is only testing symptomatic individuals. If one is symptomatic, does one get tested→ if tested, is one positive for the virus→ if positive, does one die. Hierarchical models are built at the provincial level and aggregated federally. There is little evidence in the current discourse of such an approach, provincial

---

[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] See https://www.nobelprize.org/prizes/economic-sciences/2009/ostrom/facts/
[3] On 7 April 2020, see 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; Scénarios: évolution de la COVID19, 7 avril 2020, Québec.

modeling and federal modeling appears separate.[4] Below a figure compares 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, i.e. no indication of what causal factors are included.

Figure 1: Morbidity rates, Canada (orange)/QC (blue)
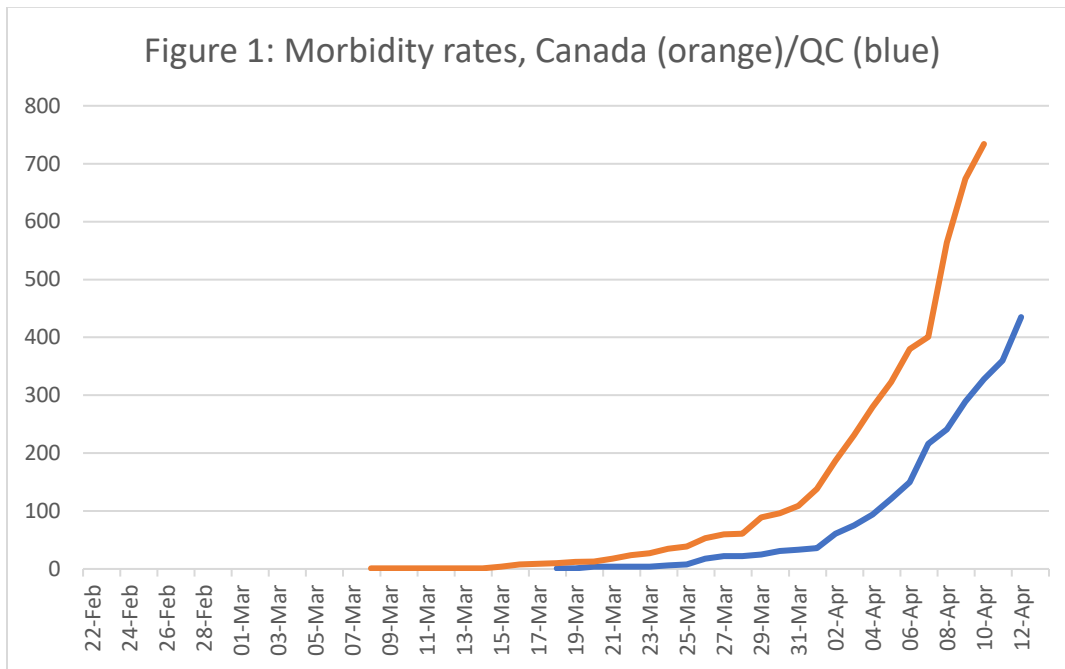
Second, the federal modeling provides no special social indications for Quebec. At no point has the typical greeting of cheek-to-cheek kisses been identified as a vector for the social spread in Montreal and eastern areas to explain the overspread in Quebec and the transmission at the ON/QC border areas. 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-powerful[5].

As a data scientist and collective action bargaining (failure) scholar teaching a graduate course on international institutions including 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 increase the cleavage. Then Italy, got sick.

When Italy got sick, discussions shifted due to family connections in Europe for many. As Sandler's book required for the course, we discussed 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 globally with a diversity of undergraduate degrees) discussed this

---

[4] Federal modeling document shows cases starting from 15 March, though empirically untrue as above. See "COVID-19 in Canada," by Public Health Canada on 9 April 2020, slide 7. Moreover, the slope of the orange line does not transfer completely upon the line "shown" in slide 11.

[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. From 7 February 2020.

challenge[7], all of the various factors to put in a predictive model of (international/national collective action success/failure) and we uniformly felt, the 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 Dashboard[9]. 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 "predictions" served as hypotheses (estimates). And, as a data scientist, I "saw" the observed and "finished the function mentally" before mathematically. Armed with an "expert" understanding, I reversed the math on the low prediction (1263) in a few hours.[10]

Despite an intellectual win, the loss is the public put faith and credibility into the government's prediction. The QC government was understandably reticent to offer predictions, given the asymmetric information experts held. In other words, the impossibility to put together the factors one requires defaulted the government into offering a "simplistic" prediction, I discovered to non-surprise one can 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 day[11]. Additionally, a linear prediction is not the appropriate functional form of the relationship, i.e. exponential.

Figure 2 visualizes the equation for a line (Y = aX + b) or Expected QC mortality rate[12] = (time*46.355) plus 150.

Figure 2: Quebec's low prediction
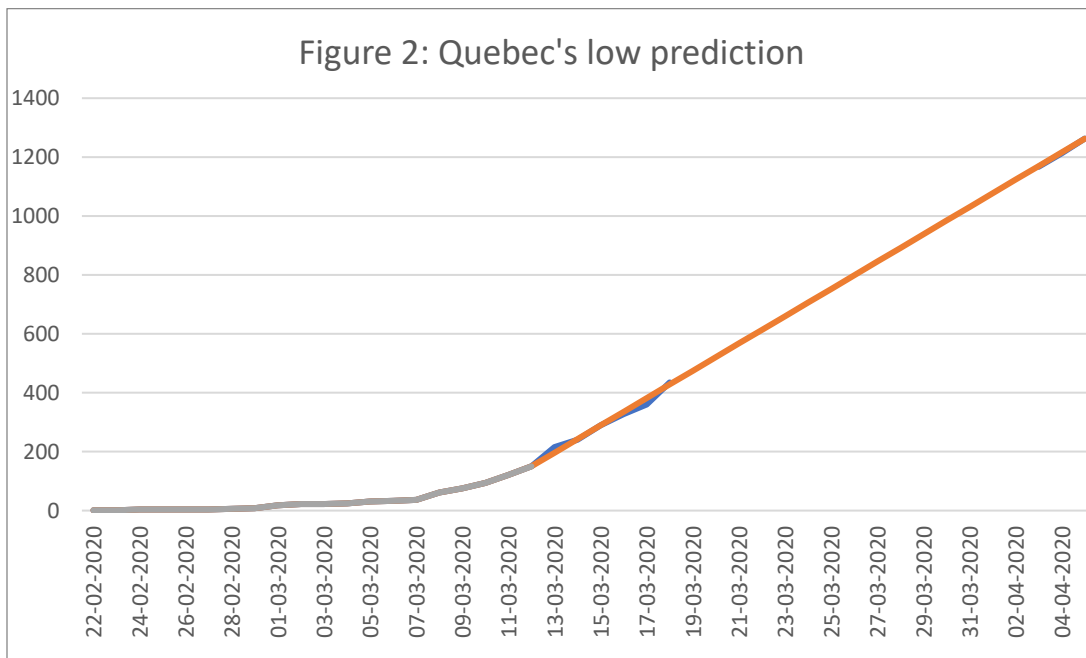
A realization more unsurprising knowing how little data science training is required in Canadian in political science departments[13]. Again, the process sequencing and delay to morbidity suggests the last week's data (the blue line above starting at 150), possibly indicates a world without maximal social distancing interventions. If social behavior is slowing the spread to and preventing mortality, then the data should start to flatten out in the 7-10 days or a virus cycle.[14]

---

[7] Across 7 course meetings of 1.5 hours over 5 weeks.

[8] See footnote 3

[9] See, https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

[10] Public tweet 8 April 2020 from author; data available upon request.

[11] grey line

[12] orange line

[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 from the first Canadian death in BC followed by the first QC death.

My third shift the next evening was occupied by reversing the higher prediction 8860 completed by fitting a scenario equal to the logic of "7.5-8 times worse". Expected QC mortality rate High prediction (blue line in figure 3 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 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 failed to present the information credibly and clearly to the public, a story not to omit.

Figure 3: QC MR predictions, low (orange) vs. high (blue)
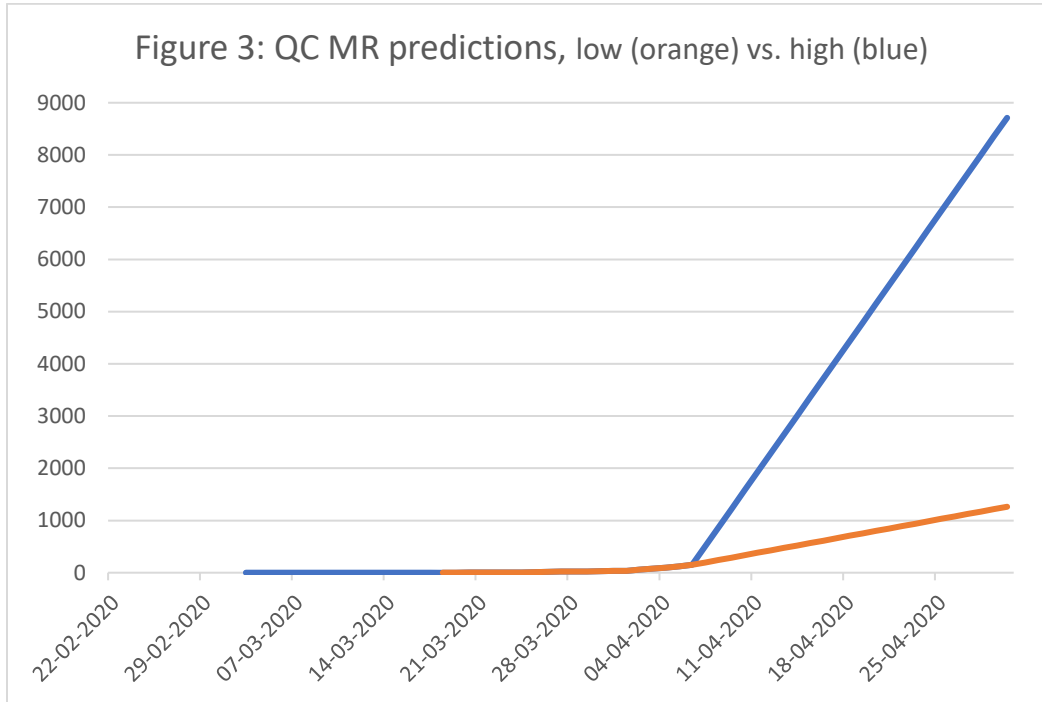
As a data scientist, economist, and intellectual, to be relevant any public policy inquiry into the data and morbidity rates offer a relevant alternative if the evidence presented confirms a lack of science. And, in particular, survival models of data analysis[15] mobilize a distinct vocabulary discussing the effects of interventions (human/medical) on the proportional hazard (i.e. differing rate) of death. In (cancer) survival models, there is attention to the differential effects of interventions across patients due to pre-existing conditions, (medically or genetically). Understanding the metaphors and language of 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 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] It permits identification of the deaths flattening by proxy, because the rate of change will be zero or negative. In Figure four below: Best: rate is consistently negative & increasingly negative

Good: rate is non-zero & negative   Neutral: rate is at zero (no growth)

Bad: rate is non-zero & positive   Worst: rate is consistently positive & growing

---

[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.*
[16] It can be adjusted for population density.
[17] Public tweet from author, 8 April 2020.

Figure 4: Comparative death rate deltas
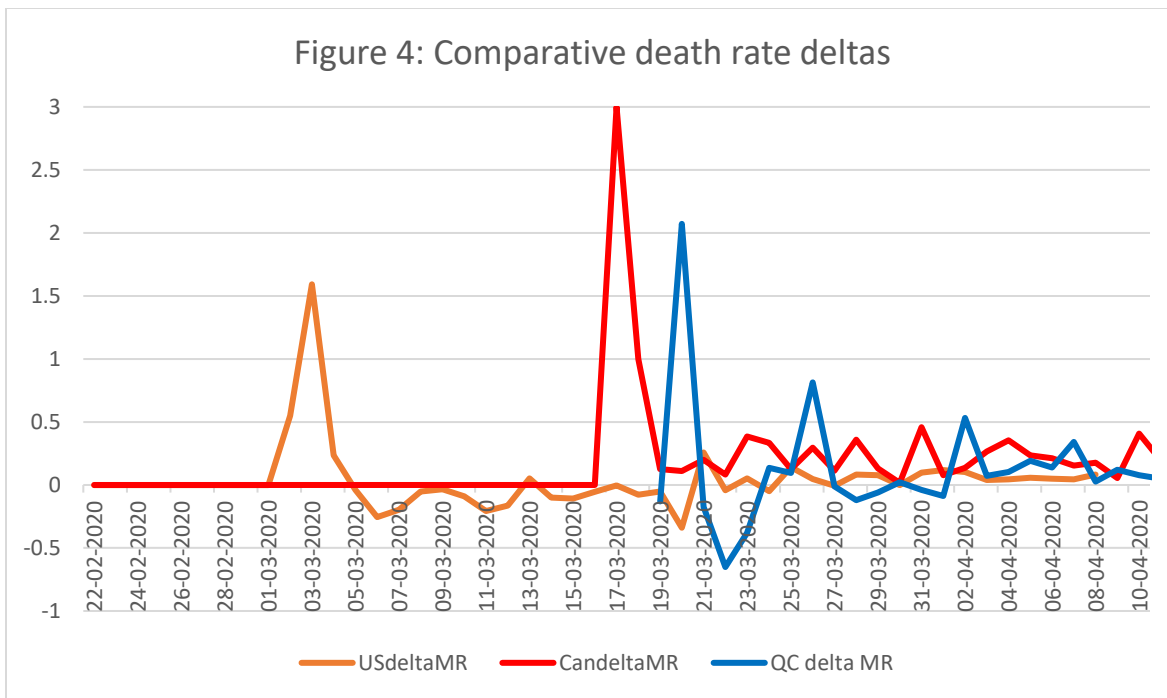
Predicting the total deaths federally requires massive data and information, which is nearly impossible, yet one can collect evidence indicating tendencies to inform policy making. Current world: Bad signs though QC is moving closer to zero showing signs of neutral[18]. I am not alone in skepticism about 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: models[19] for various morbidity predictions & data must be available, the discourse should shift to a regional/local strategy for containment and collective action, all humans must accept the short-term costs/inconveniences so medium to long term (limited) restrictions can be relaxed; vigilance is required for a year or longer until the public good (the vaccine) is made available[20]. 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. Over coming a looming collective action failure happens when individuals value society more than their utilitarian interest. It translates to remaining physically distant while socially together and Canadian geography indicates that is what this nation is built upon.

---

[18] Montreal is distinct among Quebec data.

[19] Programming code and datasets could alternatively be made public for data scientists.

[20] Sandler also offers several collective action recommendations to increase the chances a vaccine is developed (2004).