Journal of Pharmaceutical Care & Health Systems

OPEN ACCESS Freely available online

Mini Review

Public Health Policymaking using Insights from COVID-19 Modelling with News Sentiment

Ioannis Chalkiadakis[1*], Kevin Hongxuan Yan[2], Gareth W. Peters[3], Pavel V. Shevchenko[4]

[1]School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, Scotland, UK;[2]Department of Chinese Academy of Science, Beijing, China;[3]Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Scotland, UK; [4]Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, Australia

ABSTRACT

Throughout the COVID-19 pandemic, severe contact restriction measures and social mobility limitations had to be put in effect by governments all over the world, to limit the exposure of the population to the novel coronavirus. These public health policy decisions were made by considering the output of statistical models for infection rates in national populations. To assist in the public effort, we conducted research where we modelled the temporal evolution of national-level infection counts for the United Kingdom (UK-Wales, England, Scotland), Germany, Italy, Spain, Japan, Australia and the United States for the period January 2020 to January 2021, in order to better understand the most reliable model structure for the COVID-19 epidemic growth curve. We achieved this by exploring a variety of stochastic population growth models and comparing their calibration, with and without an exposure adjustment, to the most widely used growth rate model, the Gompertz population model, often referred to in the public health policy discourse over the past year.

In this work, we explore the statistical concept of model risk, which manifests in the inability to adequately capture the behavior of the disease progression growth rate curve. Model risk is mathematically characterized as having two components: The dispersion of the observation distribution and the structure of a function over time that describes the force of infection (the "intensity function"). Furthermore, we investigated how to include in these population models the effect that governmental interventions have had on the number of infected people. This was achieved through the development of an exposure adjustment to the force of infection comprised of a tailored sentiment index constructed from various authoritative public health news reporting, namely major global circulation newspapers, including The New York Times, The Guardian, The Telegraph, and Reuters global blog, as well as international, acknowledged health authorities, i.e. the European Centre for Disease Prevention and Control, the United Nations Economic Commission for Europe, the United States Centers for Disease Control and Prevention, and the World Health Organization.

Our research revealed that the baseline Gompertz model is unable to adequately capture the pandemic evolution for all countries throughout the period of study, and, in addition, models that incorporated the proposed sentiment adjustment are better able to calibrate to the infection spread in all countries under study, particularly during the early stages of the pandemic.

Keywords: COVID-19; GLARMA; Growth models; Model risk; Natural language processing; Sentiment analysis

Abbreviations: NYT: New York Times; ECDC: European Centre for Disease Prevention and Control; USCDC: Disease Control and Prevention; WHO: World Health Organization; SDE: Stochastic Differential Equation.

INTRODUCTION

A newly surfaced coronavirus strain has led, over the past year, to the onset of a global devastating pandemic that still continues into 2021, and which became widely known as the COVID-19 respiratory disease. COVID-19 has had an immense impact on society in multiple ways, including significant mortality and morbidity, long term health effects (long-COVID) and significant impact on the economies globally. Therefore, joining the global battle against COVID-19, we considered crucial that we study the statistical properties of the evolution of this disease so that we address statistical questions such as "Why were so many disease

Correspondence to: Ioannis Chalkiadakis, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, Scotland, UK, E-mail: ic14@hw.ac.uk

Received: August 10, 2021, **Accepted:** August 24, 2021, **Published:** August 31, 2021

Citation: Chalkiadakis I, Yan KH, Peters GW, Shevchenko PV (2021) Public Health Policymaking using Insights from COVID-19 Modelling with News Sentiment. J Pharma Care Health Sys S7:237.

growth rate projections so significantly wrong in the early stages of the pandemic?"

LITERATURE REVIEW

As scientists feeling the need to support our community and to contribute to the fight against the pandemic, we investigated this question from a statistical perspective based on a statistical analysis of model risk, to assist policymakers and public health officials in managing the severe consequences of the pandemic and of the imposed countermeasures. We also aimed to obtain a greater understanding of how the communication of public health announcements affected the populations' behavior, and if this had a noticeable effect on "flattening the curve" of infection cases; in other words, whether governments successfully communicated how critical it is that the public follow the protective policies taken. We quantified this effect through daily changes in infection rates as a result of public health policy and information announcements, whose impact we extracted over time *via* text mining methodologies from a variety of press releases and news articles from authoritative news agencies and public health authorities that included, among others, the New York Times (NYT), the Guardian, the Telegraph, the European Centre for Disease Prevention and Control (ECDC), the US Centre for Disease Control and Prevention (USCDC) and the World Health Organization (WHO).

Our research results gave us significant insight on two additional questions, namely "What is the most reliable and accurate way to build an epidemic growth model for this disease?" and "Can one assess the influence of public policy and public health reporting on the dynamics of the COVID-19 pandemic spread over time?"

We need to remark that there are many ways to model epidemics, through for instance compartmental epidemic models that capture individuals in groups of stages of, for example, healthy, infected, recovered or deceased [1]. These models are primarily useful for a detailed epidemiological analysis and analysis of vaccine response, however, they often utilize a few key parameters in the estimation that strongly affect the model outputs, such as the reproductive number which has been hotly debated in the epidemiological studies of COVID-19 [2]. Alternative approaches to epidemic modelling include stochastic epidemic models that focus more on models based on a Stochastic Differential Equation (S.D.E.) formulation, usually under certain assumptions to obtain tractability, which, however, may be too simple to capture the pandemic evolution effectively [3].

Given all the different modelling approaches, it is therefore critical to understand that each modelling approach is suitable for specific purposes, and although the aforementioned models are required for sophisticated epidemiological disease modelling at the individual demographic level or to better understand vaccine program responses, they are not the models that public health policymakers resort to. Often, throughout the COVID-19 pandemic crisis, the public health decision-makers in governments had to abstract from such detailed models and instead focus on simpler infection growth rate models that are better suited to national-level epidemiological analysis. This class of models is the one we focus on in our analysis. The most popular of these models used during the COVID-19 pandemic was the Gompertz model, as discussed in which forms the baseline in our analysis [4].

The evolution of the pandemic may be captured through a number of different proxies which form the datasets that researchers can access. These include time-series of daily death counts, new infections, total cumulative infections, number of hospitalizations, number of patients in Intensive Care Units, virus presence levels in public sewers etc. We have selected to study the daily number of infections at the national level, as the number of deaths is often under-reported or prone to reporting errors, as it is not only based on the exact daily counts but also depends on the governmental reporting policy. This is often affected by overloaded hospital and coroner reporting systems during the pandemic, hence making the use of this data less reliable for the models we want to calibrate.

In addition, we rely on data aggregated at the national level since we seek the policymakers' perspective for the prediction of the infection spread at a large-scale population level. Our research contributes to the epidemiological growth rate model literature by extending the basic macro infection growth rate models in four major ways: i) we develop advanced, flexible time-series models that incorporate stochastic population growth models; ii) we modify our observation distribution to better capture the different phases of COVID-19 spread in the community; iii) we develop a Bayesian estimation procedure for these models that readily offers uncertainty quantification; iv) we incorporate a novel exposure adjustment feature that involves a custom public health news announcement sentiment index. This feature allows public health officials to quantify statistically the impact of public health announcements and news releases in terms of their influence on the public's behavior, as reflected by daily changes in infections as people learned more about COVID-19 through health and news releases.

The incorporation of public sentiment information in the models allows us to obtain an insight into the effectiveness of public information and health campaigns. Our method constitutes a novel way to measure how well the pandemic protective measures are received by the public, and whether or not people adhere to the measures taken. These concepts are inherently included in the number of infected cases, and are therefore very useful to incorporate in epidemic models, as has been demonstrated in related works [5,6]. Moreover, our results demonstrate that textual data obtained from public news is a suitable proxy to public sentiment, owing to the effect of the pandemic on linguistic expression in the time of COVID-19 [7-12].

DISCUSSION AND CONCLUSION

The seven countries we analyzed were selected due to the varying epidemic spread profiles (United Kingdom, Germany, Spain, Italy, United States, Japan, Australia) they exhibit. We set up our analysis to include pre- and post-vaccination phases, and showed that the baseline Gompertz model is not able to adequately describe data that span both phases, due to the fact that the COVID-19 pandemic exhibited significantly different epidemiological profiles in the two periods under study, such as, in particular, the rapid growth rate in the second wave of the pandemic starting in Autumn 2020. These results make the model risk component explicitly clear, and illustrate that there may be significant repercussions if policy-makers are not aware of it.

Therefore, policy-makers ought to pay particular attention to the characteristics of the growth curves of the number of infections, and if necessary, employ more advanced population growth models such that they include the identified curve features. In this way, model risk will be reduced, provided also those decisions on the selected models are reviewed frequently while continuously incorporating new data that express the pandemic's evolution. Furthermore, we contributed a novel sentiment index that was obtained from news articles and institutional announcements about the COVID-19 pandemic, as communicated by institutions and national Centres for Disease Control. The inclusion of the sentiment index in the population growth models *via* an exposure adjustment, revealed that at the onset of the pandemic the in-sample model fit is much better, namely the sentiment adjustment significantly helps the model capture the growth rate of infected cases. This is especially important for model assessment, and assessment of the effectiveness of the applied pandemic countermeasures, protective policies, and the way they were communicated to the public. Our results prove that this research work is particularly useful for designing, communicating and evaluating protective policies during extreme events, as well as scenario generation to better prepare for crises in the future.

REFERENCES

1. ChenYC, Lu PE, Chang CS, Liu TH. A time-dependent SIR model for COVID-19 with undetectable infected persons. IEEE Transactions Network Sci Eng 2020;7(4): 3279-3294.
2. Sunhwa C. Estimating the reproductive number and the outbreak size of COVID-19 in korea Epidemiol Health 2020;42(0): e2020011.
3. Platen E. Stochastic modelling of the COVID-19 epidemic. SSRN 2020;3586208.
4. Wüthrich MV. Corona COVID-19 analysis: Switzerland and Europe. SSRN 2020;04.
5. Broniatowski DA, MJ Paul, M. Dredze. National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza Epidemic. PLOS One 2013;8(12).
6. Joshi A, Sparks A, McHugh J, Karimi S, Paris C, C R MacIntyre. Harnessing tweets for early detection of an acute disease event Epidemiol 2020;31(1): 90-97.
7. Oxford English Dictionary Editorial. Corpus analysis of the language of COVID-19 Oxford English Dictionary blog 2020;04.
8. Chalkiadakis I, Yan H, Peters GW, Shevchenko PV. Infection rate models for COVID-19: Model risk and public health news sentiment exposure adjustments. PLOS One 2021;16(6): 1-39.
9. He S, Peng Y, Sun K. SEIR modeling of the COVID-19 and its dynamics. Nonlin Dynam 2020;101(3): 1667-1680
10. Liu Z, Magal P, Seydi O, Webb O. A COVID-19 epidemic model with latency period. Infect Disea Model 2020;5: 323-337.
11. Paton B. Social change and linguistic change: The language of COVID-19 Oxford English Dictionary blog 2020;04.
12. Ostermann M. Joannidis M. Acute kidney injury 2016: Diagnosis and diagnostic workup. Crit Care 2016;20(1): 299.