Author/s:
Moss, R

Title:
Commentary on “Transparent modeling of influenza incidence”: Because the model said so

Date:
2022-04-01

Citation:
Moss, R. (2022). Commentary on “Transparent modeling of influenza incidence”: Because the model said so. International Journal of Forecasting, 38 (2), pp.620-621. https://doi.org/10.1016/j.ijforecast.2021.01.028.

Persistent Link:
https://hdl.handle.net/11343/301555

# Commentary on “Transparent modeling of influenza incidence”: Because the model said so.

## Introduction

With limited data and prior experience to guide national responses to emerging infectious diseases such as COVID-19, models provide a means of assessing the likely impact of unmitigated outbreaks and exploring the relative merits of different response options. This is but one example of how models are being used with increasing frequency to inform policy decisions that affect us all. If we are to trust such decisions, the justification should be “the model said so, because . . . ”, not “because the model said so”. As one example, having clear rationale to justify measures put in place in Australia to mitigate COVID-19 was crucial in maintaining public trust and promoting compliance with these measures (Leask & Hooker, 2020; Seale et al., 2020).

The emergence of “big data” and machine learning has resulted in increasingly complex models, which presents a very real challenge when seeking to explain why a model produces the results that it does. This has stimulated the development of “explainable artificial intelligence” (XAI) methods, which aim to allow humans to understand these models and to determine when they should, and should not, be trusted (Samek et al., 2017). In contrast, simple models are much easier for a human to understand, and are capable of matching, or even out-performing, complex models (Sherden, 1997).

Katsikopoulos et al. (this issue) advocate using simple heuristics as a baseline against which the performance of complex models should be evaluated, and argue that using simple rules can yield algorithms that are both accurate and understandable. They propose a simple “recency heuristic” for influenza-like illness — this week’s data will be the same as last week’s data — and pit this against Google Flu Trends, an extremely complex model built from huge quantities of data. To compare such models for the purpose of informing real-world decisions, several questions must be addressed: what do we need from the model, does the model meet this need, and can we understand why it does/does not?

## Model purpose

What quantities do we need to predict, and when is a prediction *good enough* to inform our decision? A model should only be one part of a decision support system that recognises the broader social and political context (Shearer et al., 2020). A perfect prediction is rarely (if ever) required, let alone possible, and we may instead ask: how much better/worse is each model than the other candidates? A simple model can then be a useful baseline against which to evaluate other models, as Katsikopoulos et al. demonstrate here, and as used in the US CDC FluSight competition (Lutz et al., 2019). Decisions may also involve actions that modify the very process we are trying to predict, such as deciding whether to make facemasks compulsory to help reduce the spread of COVID-19. This requires the model to make predictions for two different scenarios: one where facemasks are worn by most people, and another where they are not.

## Model suitability

Is the model capable of providing *useful* predictions, which perform well against our chosen targets and can support decision-making? In the context of infectious disease outbreaks, the prediction lead time is critical to informing preparedness and response activities. The one-week lead time of the recency heuristic is too short; a more complex model is required and, as Katsikopoulos et al. indicate, longer-term predictions of influenza activity are much more difficult (Wilke & Bergstrom, 2020). One alternative is the SIR model family, which is also based on simple rules. They can be used to make predictions with a sufficient lead time to inform infectious disease responses (Yang et al., 2015; Moss et al., 2019; Kramer et al., 2020) and to evaluate the impact of different decisions (Moss et al., 2020). However, fitting SIR-type models to available data can be extremely challenging. In the absence of a "perfect" model, we can capitalise on the relative strengths of different models by combining them into an ensemble that can out-perform each individual model (Reich et al., 2019; Chowell et al., 2020).

## Model transparency

Can we examine the model and identify under what circumstances we should trust it? We need models that make good predictions for the *right* reasons. As model complexity increases, models can become harder to understand and may "over-fit" to irrelevant features of the data on which they are trained; as a result they may perform badly on new data. But simple, easy-to-understand rationales can also lead us to draw erroneous conclusions! Consider the paradox of moderate infection control: for pathogens that cause more severe disease in older persons, limiting transmission can *increase* the burden of disease (Cohen & Lipsitch, 2008). Combining models in an ensemble can mitigate the limitations of each individual model, but may also reduce the transparency of the results.

## Conclusion

When decisions are influenced by predictive models, we must understand the strengths and limitations of those models. To be useful in this context, a complex model should out-perform simple models and be transparent in doing so. To paraphrase George Box, all models are wrong but some are useful *if* we can verify that they yield sufficiently-good predictions, and we can understand why. A model that cannot be explained should not inform decisions that require justification.

## Acknowledgements

Robert Moss is supported by two National Health and Medical Research Council (NHMRC) Centres of Research Excellence: The Australian Partnership for Preparedness Research on Infectious Disease Emergencies (APPRISE, GNT1116530) as an APPRISE Research Fellow; and Supporting Participatory Evidence use for the Control of Transmissible Diseases in our Region Using Modelling (SPECTRUM, GNT1170960).

# References

Chowell, G., Luo, R., Sun, K., Roosa, K. M., Tariq, A., & Viboud, C. G. (2020). Real-time forecasting of epidemic trajectories using computational dynamic ensembles. *Epidemics*, 30, 100379.

Cohen, T. & Lipsitch, M. (2008). Too little of a good thing: A paradox of moderate infection control. *Epidemiology*, 18(4), 588–589.

Katsikopoulos, K., Şimşek, Ö., Buckmann, M., & Gigerenzer, G. (2021). Transparent modeling of influenza incidence: Big data or a single data point from psychological theory? *International Journal of Forecasting*.

Kramer, S. C., Pei, S., & Shaman, J. (2020). Forecasting influenza in Europe using a metapopulation model incorporating cross-border commuting and air travel. *PLOS Computational Biology*, 16(10), e1008233.

Leask, J. & Hooker, C. (2020). How risk communication could have reduced controversy about school closures in Australia during the COVID-19 pandemic. *Public Health Research & Practice*, 30(2), e3022007.

Lutz, C. S., Huynh, M. P., Schroeder, M., Anyatonwu, S., Dahlgren, F. S., Danyluk, G., Fernandez, D., Greene, S. K., Kipshidze, N., Liu, L., Mgbere, O., McHugh, L. A., Myers, J. F., Siniscalchi, A., Sullivan, A. D., West, N., Johansson, M. A., & Biggerstaff, M. (2019). Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. *BMC Public Health*, 19, 1659.

Moss, R., Wood, J., Brown, D., Shearer, F. M., Black, A. J., Glass, K., Cheng, A. C., McCaw, J. M., & McVernon, J. (2020). Coronavirus disease model to inform transmission reducing measures and health system preparedness, Australia. *Emerg Infect Dis*, 26(12).

Moss, R., Zarebski, A. E., Carlson, S. J., & McCaw, J. M. (2019). Accounting for healthcare-seeking behaviours and testing practices in real-time influenza forecasts. *Tropical Medicine and Infectious Disease*, 4, 12.

Reich, N. G., McGowan, C. J., Yamana, T. K., Tushar, A., Ray, E. L., Osthus, D., Kandula, S., Brooks, L. C., Crawford-Crudell, W., Gibson, G. C., Moore, E., Silva, R., Biggerstaff, M., Johansson, M. A., Rosenfeld, R., & Shaman, J. (2019). Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. *PLOS Computational Biology*, 15(11), e1007486.

Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable Artificial Intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv e-prints*, 1708.08296.

Seale, H., Heywood, A. E., Leask, J., Sheel, M., Thomas, S., Durrheim, D. N., Bolsewicz, K., & Kaur, R. (2020). COVID-19 is rapidly changing: Examining public perceptions and behaviors in response to this evolving pandemic. *PLOS ONE*, 15(6), e0235112.

Shearer, F. M., Moss, R., McVernon, J., Ross, J. V., & McCaw, J. M. (2020). Infectious disease pandemic planning and response: Incorporating decision analysis. *PLOS Medicine*, 17(1), e1003018.

Sherden, W. A. (1997). *The Fortune Sellers: The Big Business of Buying and Selling Predictions.* Wiley.

Wilke, C. O. & Bergstrom, C. T. (2020). Predicting an epidemic trajectory is difficult. *Proceedings of the National Academy of Sciences*, 117(46), 28549–28551.

Yang, W., Cowling, B. J., Lau, E. H. Y., & Shaman, J. (2015). Forecasting influenza epidemics in Hong Kong. *PLOS Computational Biology*, 11(7), e1004383.