

August 20, 2023

Dear Editors,

**A simulation-based approach for estimating the time-dependent reproduction number from temporally aggregated disease incidence time series data**

**I Ogi-Gittins, WS Hart, J Song, RK Nash, J Polonsky, A Cori, EM Hill, RN Thompson**

We would like to submit the above manuscript for consideration for publication in *Epidemics*. All authors have read the final manuscript and approve of its submission in the present form.

During infectious disease outbreaks, policy advisors use assessments of pathogen transmissibility to guide public health measures. A commonly used metric is the time-dependent reproduction number, (informally referred to as the R number). If is above one, then its precise value determines the proportion of transmissions that must be prevented to bring an outbreak under control.

However, estimation of is beset by temporal aggregation of outbreak data. For example, weekly case numbers are often reported, rather than daily case numbers. As we show, estimates obtained using standard inference methods can be incorrect if the timescale of transmission is shorter than the timescale of data recording; for example, if an infected individual can transmit a pathogen within a few days of becoming infected (as occurs for pathogens such as influenza viruses) but cases are only reported weekly.

In this manuscript, we present a novel simulation-based modelling framework for estimating from temporally aggregated disease incidence time series data. We compare estimates obtained using this novel approach with corresponding estimates from a frequently used inference method (the Cori method, which underlies the R software package EpiEstim). We first use a simulated outbreak dataset to demonstrate that the simulation-based method provides accurate estimates in a scenario in which the true value of is known. We then go on to apply our method to two outbreak datasets for influenza in 2019-20 and 2022-23 in Wales (in the United Kingdom). As we show, our approach is easy to use (it simply requires repeated model simulation) and leads to more accurate estimates of than commonly used methods when disease incidence time series data are temporally aggregated.

We expect the development and application of our method for estimating to be of interest not only to mathematical modellers, but also to epidemiologists and public health policy specialists. Our work has clear implications for the control of infectious disease outbreaks, and we believe that it will appeal to the wide readership of *Epidemics*.

Thank you for considering our manuscript.

Yours faithfully, on behalf of all authors,

Robin N. Thompson