engineering proceedings

MDPI

Proceeding Paper

Estimation and Implication of Time-Varying Reproduction Numbers during the COVID-19 Pandemic in the UK †

Jiangjiang Yan [1], Ruochen Huang [2,*] and Wuliang Yin [1]

[1] Department of Electrical and Electronic Engineering, University of Manchester, Manchester M13 9PL, UK; jiangjiang.yan@postgrad.manchester.ac.uk (J.Y.); wuliang.yin@manchester.ac.uk (W.Y.)
[2] College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China
* Correspondence: ruochen_huang@fzu.edu.cn
† Presented at the IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability, Tainan, Taiwan, 2–4 June 2023.

Abstract: Infectious illness prevention and control is an important part of public health management. The early monitoring and numerical modelling of incidence data can help with the efficient prevention and control of infectious disease prevalence. The reproduction number R, as an essential index to understand the dynamics of COVID-19, can be predicted by using confirmed new incidence cases and serial interval data in the datasets provided by UK government. In this paper, an extended model is proposed to account for variable reporting rates instead of 1 for the estimation of the R number. The impact of using various modelling parameters is also evaluated, which provides insight into how to select a set of suitable parameters in modelling. The variation of the estimation of the R number by incorporating variable reporting rates can be observed and assessed.

Keywords: reproduction number; reporting rates; incidence cases; serial interval distribution; COVID-19

check for updates

Citation: Yan, J.; Huang, R.; Yin, W. Estimation and Implication of Time-Varying Reproduction Numbers during the COVID-19 Pandemic in the UK. *Eng. Proc.* **2023**, *55*, 20. https://doi.org/10.3390/engproc2023055020

Academic Editors: Teen-Hang Meen, Kuei-Shu Hsu and Cheng-Fu Yang

Published: 29 November 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Infectious illness prevention and control are an important part of public health and safety, as they have an impact on human health and social stability. Large-scale epidemics are likely to emerge if infectious disease propagation is not successfully controlled. During COVID-19, the importance of epidemic monitoring, modelling, and understanding, as well as early warning and policy guidance, is strongly emphasised.

Simultaneously, with the rapid advancement of information technology and significant rise in computing power, the data analytics (DA)-based modelling and monitoring of infectious illnesses have become widely accessible for public health practitioners. In this context, data analytics refers to a set of procedures for converting data into interpretable and actionable information sets [1] and for producing reliable and innovative insights into the current state of the world [2]. The discovery and tracking of epidemic diseases in the healthcare business is one of the most notable examples of data analytics.

Modelling the transmission of infectious illnesses can be performed in a variety of ways. They are largely divided into two categories: deterministic and statistical techniques. Statistical approaches use existing data to develop estimators using probability theory, whereas deterministic (mathematical) approaches focus on the construction of a mechanistic model describing disease transmission. Comprehensive evaluations of the methods can be found in works such as [3–7]. Dietz and Becker summarised multiple modelling strategies for estimating the propagation of infectious disorders in [3,4]. Perone presented hybrid models, such as the neural network autoregression (NNAR) model and the autoregressive moving average (ARIMA) model, that perform better in estimating the disease trend than single models, such as the neural network autoregression (NNAR) model and the autoregressive moving average (ARIMA) model [5]. Fraser suggested the home reproduction number [6], a method for predicting the reproduction number based

on the household model. Using this strategy, it provides advice on the measures and control of families during epidemics. Godio et al., who combined the generalised SEIR epidemiological model with swarm intelligence, linked the model to reality movement in Italy and demonstrated that the stochastic approach is better for trend estimation [8].

The estimation of the reproductive number R is particularly important among the investigations. R is the average number of secondary cases induced by a single infected person [9]. R is used to measure the transmission capability of pathogens during epidemics; therefore, its estimation is critical in the prevention and control of infectious illnesses. R monitoring offers feedback on the success of control measures and the need to reinforce control [10]. It is a useful index for understanding the dynamics of infectious illnesses. Pharmaceutical interventions (PIs), such as vaccination, and non-pharmaceutical interventions (NPIs), such as lockdowns, are two types of public health control strategies that try to prevent and/or regulate transmission in the community. The purpose of intervention activities is to lower R below 1 and to be as close to 0 as possible, thereby bringing an epidemic under control. Various elements, such as the daily confirmed data, influence the accuracy of R estimation. An estimation mistake would result from the data's inaccuracy. However, there are only a few approaches that may be used to fix this type of problem.

To solve this problem, an extended model for R estimate based on variable reporting rates is proposed in this study. The proposed model can be used to more precisely analyse the evolution of the disease. In the theoretical demonstration of the model, varying reporting rates are introduced, followed by simulation outcomes due to various modelling settings. Finally, the utilisation of the reporting rate is assessed based on R number estimation.

2. The Theoretical Model for Estimating Time-Varying Instantaneous Reproduction Numbers

The term "epidemic model" describes a mathematical model used to describe infectious illnesses quantitatively. It uses quantitative research methodologies to study infectious diseases primarily in terms of growth and transmission characteristics in order to develop a model that can accurately depict the transmission dynamic characteristics of infectious diseases. Infectious disease modelling is primarily used to investigate disease transmission mechanisms, forecast future outbreaks, and assess the effectiveness of epidemic management efforts.

Researchers at Imperial College developed a novel framework and software to estimate time-varying reproduction counts during epidemics [9]. Because it provides a framework for estimating time-varying instantaneous reproduction numbers from incidence time series during epidemics, this line of models is referred to as the ETVR model in the study. Epidemiologists and public health organisations can use the ETVR model to alter public health responses in real time [11].

The calculation of the R number, according to the ETVR model, is dependent on accurate incident data (I_a^t) and the serial interval distribution (S^t). The serial interval distribution is used to approximate the infectivity profile of the infectious disease. It is common for infectious cases to be over- or under-reported. To address this problem, it is necessary to implement a reporting rate for the estimation of the R number. From here, both the reporting rate (λ^t) and the reported daily confirmed cases (I_{rp}^t) can be used to present the actual daily confirmed cases at timestep t, as expressed below:

$$I_a^t = \lambda^t I_{rp}^t \quad (1)$$

The reporting rate is considered to be constant at timestep t, but it can vary over the course of the period. The impact of different reporting rates on the R number estimation will be examined in this study.

The total infectivity of all infected individuals over a time period t can be calculated by combining the serial interval distribution and the actual confirmed cases:

$$R^t = \frac{I_a^t}{\Gamma^t} = \frac{\lambda^t I_{rp}^t}{\sum_{n=1}^{t} S^n I_a^{t-n}} \tag{2}$$

Then, by the definition of the reproduction number R [11], it can be obtained as

$$E\left(I_a^t\right) = R^t \sum_{n=1}^{t} S^n I_a^{t-n} \tag{3}$$

The time window is used to stabilise the change in the R number, producing less statistical noise, because the R number is particularly sensitive to small time steps, making it difficult to analyse the disease's transmissibility. Furthermore, it is expected that R remains constant during the sliding time window.

As a result, the R number can be more correctly calculated using this extended model based on the ETVR model and reporting rates. It can promote the use of surveillance data, analyse infectious disease incidence time series quickly, and quantify temporal changes in the transmission intensity of future epidemics.

3. The Impacts of Modelling Parameters on the Estimation of R

The reproduction number R is used to quantify transmissibility during epidemics. It varies over time due to various factors such as changes in underlying transmission mechanisms (e.g., due to seasonality or new variants), changes in contact patterns (changes in social interaction and hygienic habits), and the impact of control measures (lockdowns, etc.).

COVID-19 data were obtained from Our World in Data (https://ourworldindata.org/covid-cases, accessed on 10 November 2021), an organisation established by the Global Change Data Lab (GCDL), the Oxford Martin School, and the University of Oxford, for use in the simulation. Because of varied configuration factors, the R estimation differs. In the following sections, the effects of various modelling parameters are assessed.

3.1. Effect of Incubation Period in Serial Interval Distributions

The infectivity profile is generated using a serial interval distribution, which is an approximation of the generation time. As a result, it has a considerable impact on the R number estimation. Three serial interval distributions (S1, S2, and S3) are used to analyse their impact on the R number, as indicated in the Appendix A. In secondary cases, the symptom onset window was set to one day. Three alternative serial interval distributions represent varying levels of uncertainty about the onset of symptom within a single day.

The simulation results are shown in Figure 1 for three serial interval profiles. From S1 to S3, three serial interval profiles show an increase in generation time. When R > 1, as shown in Figure 1, an increase in generation time results in a general rise in estimated R for a given incidence profile. This is reasonable because as the generation time increases, the time delay between the infector and the infectee increases, necessitating a bigger R to produce the same number of cases down the line.

3.2. Effect of Time Window

The size of the time window used to estimate R is expected to have an impact on the results. Smaller R values result in a faster detection of transmission changes but higher statistical noise; larger R values result in more smoothing and lower statistical noise. Figure 2 depicts this observation.

Small windows, on the other hand, can result in highly volatile estimates with large credible intervals, whereas longer windows can result in smoothed estimates with narrower credible intervals. This is illustrated in Figure 2, where the R's standard deviation reduces as the temporal window length increases. In practice, the time window can be determined by the desired coefficient of variation or confidence level. A bigger time frame size is recommended for a low coefficient of variation.

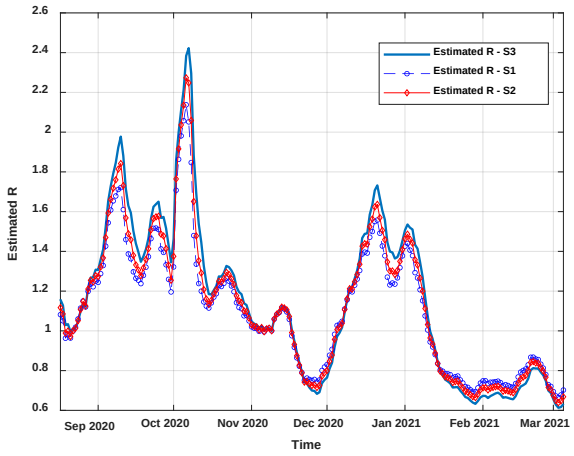

Figure 1. Three serial interval distributions and their R numbers, August 2020–March 2021.

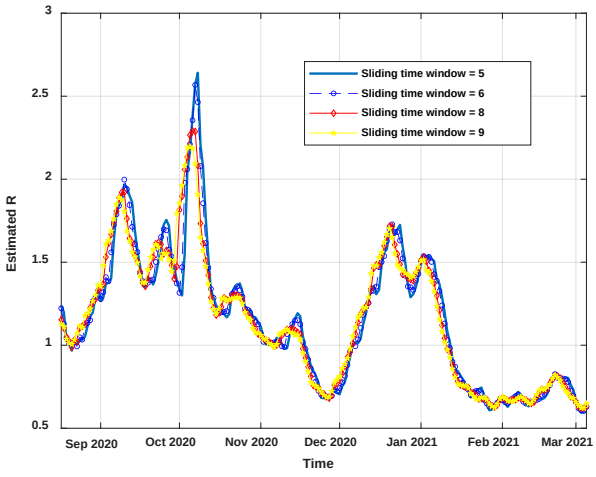

Figure 2. Estimates of R for varying time window size.

4. The Estimation of R Using the Extended Model with Variable Reporting Rates

The measurement of the R number is an important indicator for quantitatively describing infectious diseases since it could effectively reflect the transmission dynamic aspects of infectious diseases. In most research works, all cases are presumed to have been discovered and reported. However, in fact, this is not the case. As a result of including and considering the reporting rate to the gathered data, the incident data are then corrected due to over-reporting by scaling with the support of the reporting rate under 1. Because the reporting rate might be either constant or variable, both instances are covered here.

4.1. Effect of Constant Reporting Rate Due to Over-Reporting

Due to over-reporting, in this section, two over-reporting situations are simulated by keeping the reporting rate at 0.7 and 0.6, respectively. Figure 3 shows the simulated results for the estimated R.

It can be noted in Figure 3 that constant data scaling (constant over-reporting rate) has no effect on the prediction of the R number. This can be understood in the sense that the R number is a dimensionless ratio; it is not related to the overall number of incidence cases, but rather to how the cases increase or decrease, and thus is a reflection of pathogen transmission capacity during epidemics, which is the average number of secondary cases caused by an infected individual. This phenomenon is termed as the invariability of R with incidence data scaling.

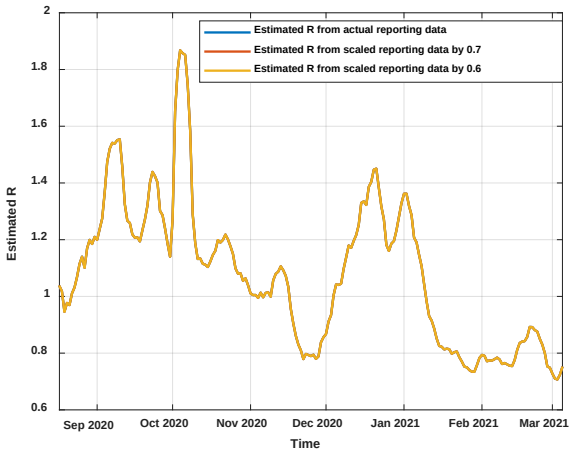

Figure 3. The results of the estimated R under constant over-reporting rates.

4.2. Effect of Variable Reporting Rates Due to Over-Reporting

Due to over-reporting, the gathered data can be adjusted by estimating the R number using the reporting rate. It cannot, however, be steady for the entire period. Herein, we present one situation whereby the reporting rate moves from a lower value to a higher value.

The new incidence data in this case are a mix of the new incidence data before the transition (scaled by 0.6) and the new incidence data after the transition (scaled by 0.7). Figure 4 shows the simulation result of the estimated R under these conditions. It appears that an erroneous R prediction occurs during the transition from a low reporting rate to a higher reporting rate (0.6 to 0.7). This can be explained by the fact that the model perceives an increase in cases as an increase in disease transmission capability because it is unaware of the variation in reporting rates.

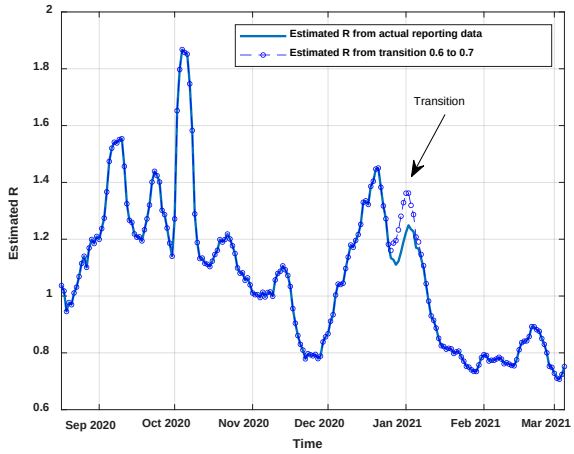

Figure 4. R numbers from the transition from a low reporting rate to a higher reporting rate (0.6 to 0.7).

5. Conclusions

For the estimation of R, we propose an extended model based on the ETVR model and reporting rate. The incidence statistics can be rectified using the reporting rate, resulting in a more accurate reproduction number for assessing infectious disease transmission, e.g., the COVID-19 pandemic.

In this paper, the mathematical manipulation for the proposed ETVR model is demonstrated. Several impacting factors were evaluated during simulations. The effect of serial interval distributions on the R number is found to be low, making their selection straightforward. The incubation period distribution has a considerable impact, with a longer incubation period leading to a higher R estimation and vice versa. Furthermore, the size of

the time window matters, i.e., smaller windows resulting in highly variable estimates with broad credible intervals.

Further, according to the simulation results from the extended model with variable reporting rates, it reflects that, as long as the fraction of asymptomatic cases and the reporting rate remain constant throughout time, the reporting rate has no significant impact on the estimation of the R number. However, when the reporting rate is variable, it has an effect on the R estimate, especially during transition times between different reporting rates.

Author Contributions: Conceptualization, R.H. and W.Y.; methodology, J.Y., R.H., and W.Y.; software, J.Y. and R.H.; validation, J.Y.; writing—original draft preparation, R.H.; writing—review and editing, W.Y.; supervision, W.Y.; funding acquisition, W.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by UK Engineering and Physical Sciences Research Council (EPSRC) with the grant number: EP/P027237/1 (title: Real-time In-line Microstructural Engineering (RIME)).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Three serial interval distributions.

S1				S2				S3			
EL	**ER**	**SL**	**SR**	**EL**	**ER**	**SL**	**SR**	**EL**	**SR**	**SL**	**SR**
0	1	3	4	0	1	4	5	0	1	5	6
0	1	4	5	0	1	5	6	0	1	6	7
0	1	5	6	0	1	6	7	0	1	7	8
0	1	5	6	0	1	6	7	0	1	7	8
0	1	6	7	0	1	7	8	0	1	8	9
0	1	6	7	0	1	7	8	0	1	8	9
0	1	6	7	0	1	7	8	0	1	8	9
0	1	6	7	0	1	7	8	0	1	8	9
0	1	7	8	0	1	8	9	0	1	9	10
0	1	7	8	0	1	8	9	0	1	9	10
0	1	7	8	0	1	8	9	0	1	9	10
0	1	8	9	0	1	9	10	0	1	10	11
0	1	8	9	0	1	9	10	0	1	10	11
0	1	9	10	0	1	10	11	0	1	11	12
0	1	10	11	0	1	11	12	0	1	12	13
0	1	11	12	0	1	12	13	0	1	13	14

Three serial interval distributions are shown here: S1, S2, and S3. EL, ER, SL, and SR are the four columns that make up the serial interval data. The lower bound of the infector's symptom onset date is EL (given as an integer). The upper bound of the infector's symptom onset date is ER (given as an integer). ER should be more than or equal to EL. The lower bound of the infected person's symptom onset date is SL (given as an integer). SR is the upper limit of the infected person's symptom onset date (given as an integer). SR should be more than or equal to SL.

References

1. Bates, D.W.; Saria, S.; Ohno-Machado, L.; Shah, A.; Escobar, G. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Aff.* **2014**, *33*, 1123–1131. [CrossRef] [PubMed]
2. Singh, M.; Bhatia, V.; Bhatia, R. Big data analytics: Solution to healthcare. In Proceedings of the 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT), Jaipur, India, 22–23 December 2017; pp. 239–241.
3. Dietz, K. The estimation of the basic reproduction number for infectious diseases. *Stat. Methods Med. Res.* **1993**, *2*, 23–41. [CrossRef]
4. Becker, N.G. *Analysis of Infectious Disease Data*, 1st ed.; Chapman & Hall/CRC Press: New York, NY, USA, 1989.
5. Fraser, C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE* **2007**, *2*, e758. [CrossRef]
6. Perone, G. Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy. *Eur. J. Health Econ.* **2021**, *23*, 917–940. [CrossRef] [PubMed]
7. Parshani, R.; Carmi, S.; Havlin, S. Epidemic threshold for the susceptible-infectious-susceptible model on random networks. *Phys. Rev. Lett.* **2010**, *104*, 258–701. [CrossRef]
8. Godio, A.; Pace, F.; Vergnano, A. SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3535. [CrossRef] [PubMed]
9. Cori, A.; Ferguson, N.M.; Fraser, C.; Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **2013**, *178*, 1505–1512. [CrossRef] [PubMed]
10. White, L.F.; Moser, C.B.; Thompson, R.N.; Marcello, P. Statistical estimation of the reproductive number from case notification data. *Am. J. Epidemiol.* **2020**, *190*, 611–620. [CrossRef] [PubMed]
11. Thompson, R.N.; Stockwin, J.E.; van Gaalen, R.D.; Polonsky, J.A.; Kamvar, Z.N.; Demarsh, P.A.; Dahlqwist, E.; Li, S.; Miguel, E.; Jombart, T.; et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* **2019**, *29*, 100356. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.