Accounting for imported cases in estimating the time-varying reproductive number of COVID-19 in Hong Kong

Tim K. Tsang[1,2], Peng Wu[1,2], Eric H. Y. Lau[1,2], Benjamin J. Cowling[1,2]

1. WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China.
2. Laboratory of Data Discovery for Health Limited, Hong Kong Science and Technology Park, New Territories, Hong Kong Special Administrative Region, China.

Corresponding author:
Tim Tsang, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 7 Sassoon Road, Pokfulam, Hong Kong Special Administrative Region, China.
Tel: +852 3917 9715; Fax: +852 3520 1945; email: timtsang@connect.hku.hk

Alternative corresponding author:
Benjamin J. Cowling, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 7 Sassoon Road, Pokfulam, Hong Kong Special Administrative Region, China.
Tel: +852 3917 6711; email: bcowling@hku.hk

Accepted Manuscript

© The Author(s) 2021. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail: journals.permissions@oup.com.

Brief summary: We developed a model to estimate separately the time-varying reproductive number (R_t) for local cases and imported cases, accounting for imperfect contact tracing of cases, and potential different infectiousness among local and imported cases.

Accepted Manuscript

ABSTRACT

Estimating the time-varying reproductive number, R_t, is critical for monitoring transmissibility of an infectious disease. The impact of imported cases on the estimation is rarely explored. We developed a model to estimate separately the R_t for local cases and imported cases, with accounting for imperfect contact tracing of cases. We applied this framework to data on COVID-19 outbreaks in Hong Kong. The estimated R_t for local cases rise above 1 in late March, 2020, which was undetected by other commonly used methods. When imported cases accounted for a considerable proportion of all cases, their impact on estimating R_t is critical.

Key word: Reproductive number, COVID-19, transmissibility, Imported cases

Accepted Manuscript

INTRODUCTION

During an emerging infectious disease outbreak, such as COVID-19, monitoring transmissibility is important for guiding the implementation of public health measures and providing situational awareness on the effectiveness of interventions [1-3]. This can be achieved by estimating the time-varying reproductive number, R_t, a measure of transmissibility of the virus over time. A wide range of methods have been proposed [3-9], especially during the period of COVID-19 outbreak when reliable methods are needed. However, the impact of imported cases on estimation of R_t is rarely explored [5].

Imported cases have not been infected locally, but could potentially generate local transmissions. The impact of imported cases on estimation of R_t would be limited when an epidemic is largely driven by infections occurring locally. However, imported cases can play a more important role in the early stages of an epidemic with few local infections, or when local transmission is being suppressed by public health measures. Some studies attempted to use information on imported cases to improve estimates of transmission [5, 10]. In particular, a study [5] proposed a modification by assuming the same infectiousness among imported and local cases, such modification may not be accurate when there are specific interventions targeting imported cases, such as quarantine for inbound travelers [11].

In some regions with effective contact tracing of cases such as Hong Kong [12], the source of infection could be determined for the majority of local cases, which provides sufficient information to estimate the transmissibility of both local cases and imported cases. However, contact tracing in reality would be imperfect and some local cases would be classified as unlinked local cases with no apparent source of infection. Accounting them in estimation of R_t would be critical to have unbiased estimates.

To investigate the impact of imported cases on estimation of transmission, we extend the framework in Cori et al [4] to estimate the R_t for local cases and imported cases separately. We develop an inference approach to estimate the R_t for local cases and imported cases, while accounting for the unlinked local cases. We used the first four months of COVID-19 outbreaks in Hong Kong to illustrate how the use of inappropriate methods to estimate R_t would lead to a biased assessment of transmissibility. We also used a simulation approach to determine the impact of unlinked local cases on our estimation approach.

Accepted Manuscript

METHODS

Sources of Data

COVID-19 has been a notifiable disease in Hong Kong since 8 January 2020. Data on laboratory-confirmed COVID-19 cases were obtained from the webpage of Centre for Health Protection. Cases are classified as "imported cases", "local cases epidemiologically linked with imported cases", "unlinked local cases" and local cases epidemiologically linked with local cases" according to their epidemiological characteristics and location of infection [12].

Statistical analysis

We examined different methods to handle imported cases in the estimation of R_t. First, when imported cases were falsely assumed as local cases, R_t could be estimated in the model described by Cori et al. [4]. In this approach, R_t was the ratio between the number of new cases at time t and the total infectiousness of cases at time t, given by $\sum_{s=1}^{t-1} I_{t-s} w_s$, where I_t was the number of new cases at time t and w_t was the probability distribution of infectiousness since infection. Second, assuming equal infectiousness for imported cases and local cases [5], the total infectiousness of cases could be represented by $\sum_{s=1}^{t-1}(I_{t-s}^{Local} + I_{t-s}^{Imported})w_s$, where I_{t-s}^{Local} and $I_{t-s}^{Imported}$ were the numbers of new local and new imported cases at time t-s, in the above model, to estimate R_t.

To account for the different infectiousness of local cases and imported cases due to travel-related measures [13], we extend above framework [4] to separately estimate R_t of local cases and imported cases. In this framework, R_t for imported cases was the ratio of the number of new local cases infected by imported cases at time t and the total infectiousness of imported case at time t, given by $\sum_{s=1}^{t-1} I_{t-s}^{Imported} w_s$. The R_t for local cases could be calculated by using the same approach (Appendix).

However, some cases are classified as unlinked local cases due to imperfect contact tracing, which can be infected from either local and imported cases. Therefore, our inference was based on a Bayesian framework and we developed a data-augmented Markov chain Monte Carlo algorithm to jointly estimate the model parameters, the probability of a local case with missing epidemiological link (p_{miss}), and the number of new unlinked cases that were actually linked with local and imported cases for each day (Appendix).

We applied this framework to the first four months of COVID-19 outbreak in Hong Kong. We could only observe epidemic curve by confirmation date but not

Accepted Manuscript

by infection date. However, given the potential for pre-symptomatic infectiousness of COVID-19 [14], it was critical to conduct analysis on epidemic curve by infection date (unobserved). Therefore, we used a deconvolution approach to reconstruct the epidemic curve by infection date from the epidemic curve by confirmation date, with the distribution of delay from infection to confirmation (Appendix). We also conducted sensitivity analyses using another deconvolution approach, and using different prior in estimation (Appendix). The probability distribution of infectiousness since infection w_t was a convolution of incubation period and the infectiousness relative to onset (allowed to be pre-symptomatic) based on viral shedding data [15]. Imported cases would not cause transmission prior to arrival, hence their infectiousness profile were adjusted (Appendix). To account for the uncertainty of input parameters such as incubation period distribution, we used an bootstrap approach (Appendix). We compared the estimates of R_t from the alternative methods. All analyses were conducted in R version 4.0.5 (R Foundation for Statistical Computing, Vienna, Austria).

Simulation study

We validated our approach and explored the impact of p_{miss} on our approach. We tested our approach for different scenarios, representing the impact on R_t from different interventions. For each scenario, we conducted simulations with different value of p_{miss}, to determine their impact on the robustness of our approach (Appendix).

RESULTS

From 23 January to 8 May 2020, there were 719 imported cases and 326 local cases detected in Hong Kong. Among the 326 local cases, there were 62 (19%), 191 (59%) and 73 (22%) cases epidemiologically linked with imported cases, with local cases, and not epidemiologically linked respectively (Figure 1A).

Based on our approach, we estimated that the surge in local case numbers corresponded to an estimated R_t for local cases greater than one during March 12 to 26, which was the period before public health measures were tightened (Figure 1B). During the period with public health measures implemented in the community, the estimated R_t for local cases was decreasing. The estimated R_t for imported cases were well below one even with there were more than 10 imported cases identified every day during early March (Figure 1C), when a 14-day quarantine was required for inbound travelers. We estimated that p_{miss} was 0.30 (95% credible interval (CrI): 0.24, 0.37). The estimated R_t was similar when

Accepted Manuscript

using another deconvolution approach to estimate infection time series (Figure S1), or when using a different prior (Figure S2).

We explored two alternative approaches to account for imported cases when estimating R_t, in which simulation studies suggested that they could be biased. In both estimations assuming the same infectiousness for local cases and imported cases (Figure 2A) [5], and assuming all cases were local cases (Figure 2B) [4], the observed local outbreaks in mid-March were missed.

Results with 200 simulated epidemics in each scenario suggested that our approach could provide unbiased estimate of R_t for local cases and imported cases in different scenarios when R_t is a constant (Scenario 1, 3) or moving from a constant to another constant (Scenario 2,4 and 5), to representing the impact of interventions, when $p_{miss} \leq 0.5$ (Figure S3-S5). However, when $p_{miss} \geq 0.6$, the estimated R_t was biased (Figure S6-7), due to insufficient information to estimate the source of unlinked local cases. There was a lag period for the estimated R_t to converge to its true value when R_t was moving from a constant to another constant. In the scenarios that R_t has stepwise changes frequently, when $p_{miss} \leq 0.5$, the estimated R_t would converge to its true value after the lag period (Figure S8). In the scenarios that R_t changed gradually, our approach could still provide the correct directions with a lag, but the exact value could be biased (Figure S9). Simulation suggested that assuming equal infectiousness would underestimate the R_t, (Figure S10) while assuming all cases were local would overestimate the R_t, during the early phase of outbreaks (Figure S11).

DISCUSSIONS

In this study, we extended the current framework for estimation of R_t to account for differential transmission from imported cases, which can separately estimate the R_t for local cases and imported cases. When control measures were targeting imported cases, our framework could estimate a lower R_t for imported cases. Also, since there would be a fraction of cases with no apparent source of infection due to imperfect contact tracing in reality, the developed estimation approach would account for this by jointly estimating the model parameters and the number of unlinked local cases that were linked with local cases or imported cases for each day during the outbreak.

We compared the extended framework with the current two approaches, namely assuming all cases were local cases, or assuming equal infectiousness for imported cases and local cases [5]. These two approaches would have little impact on measuring transmissibility in the community when the epidemic was largely driven by local transmissions, but would bias the estimated R_t in

Accepted Manuscript

measuring transmissibility in the community, when the number of local cases and imported cases were comparable. In particular, assuming all cases were local cases would overestimate the number of local transmissions as imported cases could not be infected locally. This would likely happen in the importation phase of a novel infectious disease outbreak, since local outbreaks would follow the importations, as illustrated by the outbreaks in Hong Kong.

This was improved by assuming equal infectiousness of local and imported cases [5], but might not be applicable if there were targeted control measures for inbound travelers. Once a novel infectious disease was identified, control measures targeting inbound travelers would likely to be implemented to prevent spillover to the local population, such as early March, 2020 in Hong Kong [12]. The local outbreak in late March for Hong Kong would be undetected ($R_t < 1$) if equal infectiousness for local and imported cases was falsely assumed.

Given that pre-symptomatic transmission was substantial for COVID-19, using serial intervals as a proxy of infectiousness profile would be inappropriate [4]. Hence, we used a deconvolution approach to infer the epidemic curve by infection date from the observed epidemic curve by confirmation date, using estimated infectiousness profile since infection [15]. Misspecification of the infectiousness profile since infection, or using a wrong approach to obtain the epidemic curve by infection dates such as back-shifting based on the delay distribution would bias the estimated R_t [16].

Our study has some limitations. First, we did not account for under-reporting such as mild or asymptomatic cases. Also, testing availability and criteria would also be likely changing over time. If the proportion of undetected cases were constant over the epidemic, the R_t would still be unbiased [16]. Further investigation would be necessary to develop methods to account for the changing proportion of undetected cases. Secondly, we assumed perfect classification of cases, but misspecification of the source of infection could affect the estimates (Appendix). Finally, rapid changes in the parameters could affect the estimated R_t, such as the sudden change for the delay distribution from infection to confirmation caused by overwhelmed healthcare systems.

In conclusion, we developed a methodology to estimate separately the R_t for transmissions associated with local and imported infections, accounting for the potential different infectiousness due to control measures, and imperfect contact tracing in reality. Accurate estimation of R_t allows situational awareness of transmission and better infection control in the community.

Accepted Manuscript

REFERENCES

1. Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature **2006**; 442:448-52.

2. Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. Proc Natl Acad Sci U S A **2004**; 101:6146-51.

3. Cauchemez S, Boelle PY, Thomas G, Valleron AJ. Estimating in real time the efficacy of measures to control emerging communicable diseases. Am J Epidemiol **2006**; 164:591-7.

4. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol **2013**; 178:1505-12.

5. Thompson RN, Stockwin JE, van Gaalen RD, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics **2019**; 29:100356.

6. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol **2004**; 160:509-16.

7. Bettencourt LM, Ribeiro RM. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. PLoS One **2008**; 3:e2185.

8. Scott J. A. GA, Mishra, S., Unwin J., Flaxman S. & Bhatt S. (2020). epidemia: Modeling of Epidemics using Hierarchical Bayesian Models.R package version 1.0.0 https://imperialcollegelondon.github.io/epidemia/.

9. Abbott S. HJ, Sherratt, K., Gostic K., Hickson J., Badr H., DeWitt M., Thompson R. & Funk S. (2020). EpiNow2: Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters, DOI: 10.5281/zenodo.3957489.

Accepted Manuscript

10. Fox SJ, Bellan SE, Perkins TA, Johansson MA, Meyers LA. Downgrading disease transmission risk estimates using terminal importations. PLoS Negl Trop Dis **2019**; 13:e0007395.

11. Wells CR, Sah P, Moghadas SM, et al. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. Proc Natl Acad Sci U S A **2020**; 117:7504-9.

12. Cowling BJ, Ali ST, Ng TWY, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health **2020**; 5:e279-e88.

13. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. PLoS One **2007**; 2:e758.

14. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis **2020**; 93:284-6.

15. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med **2020**; 26:672-5.

16. Gostic KM, McGough L, Baskerville EB, et al. Practical considerations for measuring the effective reproductive number, Rt. PLoS Comput Biol **2020**; 16:e1008409.

Accepted Manuscript

SOURCES OF FINANCIAL SUPPORT

This project was supported by the Health and Medical Research Fund, Food and Health Bureau, Government of the Hong Kong Special Administrative Region (grant no. COVID190118), and the Collaborative Research Fund from the University Grants Committee of Hong Kong (project no. C7123-20G).

COMPETING INTERESTS STATEMENT

BJC reports honoraria from Sanofi Pasteur, GSK, Moderna and Roche. All other authors report no other potential conflicts of interest.

ACKNOWLEDGMENTS

The authors thank Stephanie Gao, Faith Ho and Can Wang for technical assistance.

Accepted Manuscript

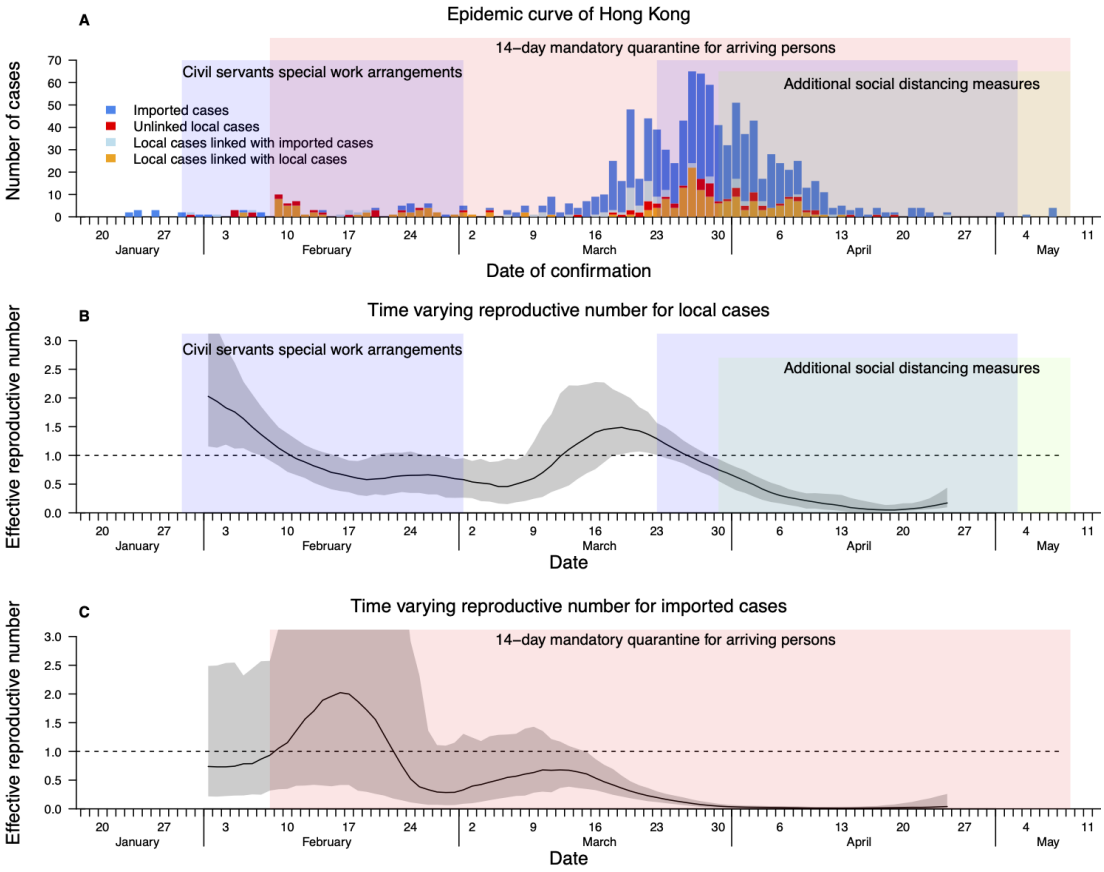
FIGURE LEGENDS

Figure 1. Epidemic curve of the COVID-19 transmission in Hong Kong from January through May 2020 (Panel A) and the estimated time-varying reproductive number for local cases (Panel B) and imported cases (Panel C). The blue-, green- and red-shaded area indicate the periods for civil servant special work arrangement, additional social distancing measures, and the 14-day mandatory quarantine for arriving persons, respectively. In panel B and C, the black lines represented the point estimate, and the black-shaded areas indicate the 95% credible intervals.

Figure 2. Comparison of estimation of time-varying reproductive numbers using our proposed framework, with alternative approaches that assuming equal infectiousness for local cases and imported cases (Panel A), or assuming all cases were local cases in the analysis (Panel B). The blue-, green-shaded area indicated the period for civil servant special work arrangement and additional social distancing measure respectively. In panel A and B, the black lines represent the point estimates, and the black-shaded areas indicate the 95% credible intervals, from our proposed method. In panel A, the blue solid and dashed line represent the point estimates, and the black-shaded areas indicate the 95% credible intervals from the method assuming the same infectiousness of local and imported cases. In panel B, the red solid and dashed line represent the point estimates, and the black-shaded areas indicate the 95% credible intervals from the method assuming all cases were local cases.

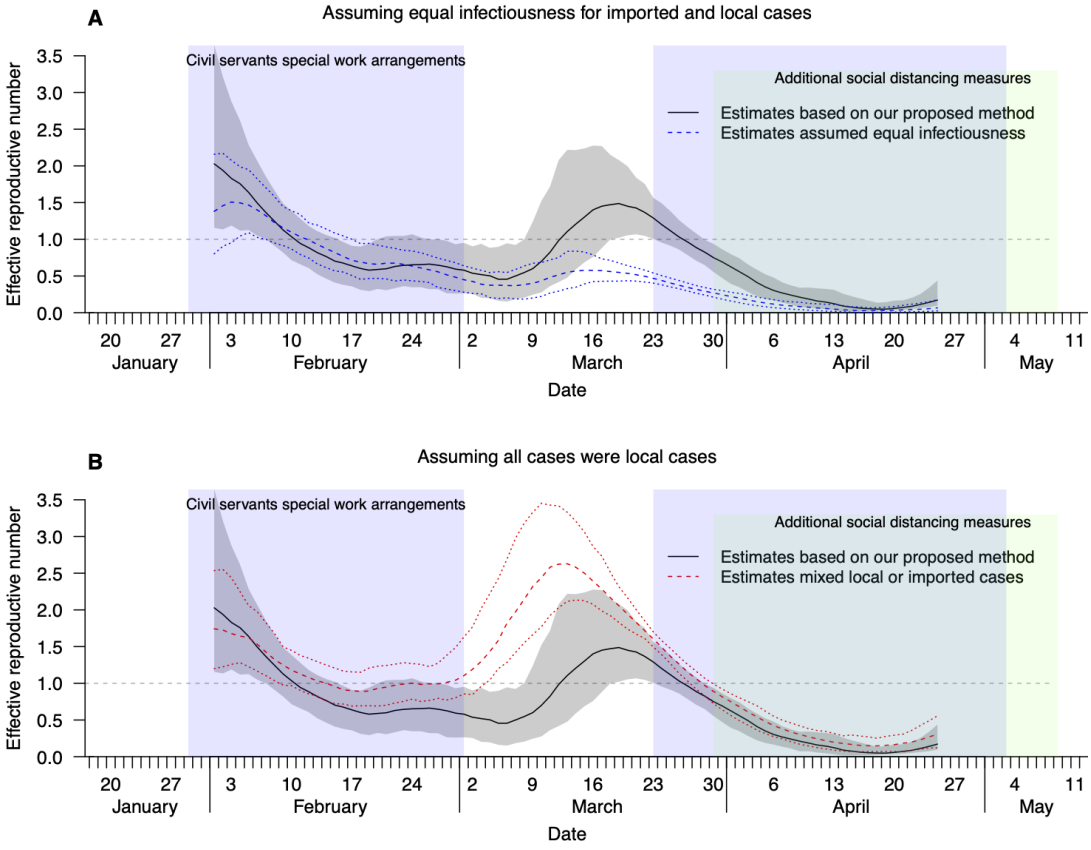
Accepted Manuscript

Figure 1



Accepted

Figure 2



Accepted