

The effect of variation of individual infectiousness on SARS-CoV-2 transmission in households

Tim K Tsang^{1,2*}, Xiaotong Huang¹, Can Wang¹, Sijie Chen¹, Bingyi Yang¹, Simon Cauchemez^{3†}, Benjamin John Cowling^{2**†}

¹WHO 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, China; ²Laboratory of Data Discovery for Health, Hong Kong, China; ³Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, Paris, France

Abstract Quantifying variation of individual infectiousness is critical to inform disease control. Previous studies reported substantial heterogeneity in transmission of many infectious diseases including SARS-CoV-2. However, those results are difficult to interpret since the number of contacts is rarely considered in such approaches. Here, we analyze data from 17 SARS-CoV-2 household transmission studies conducted in periods dominated by ancestral strains, in which the number of contacts was known. By fitting individual-based household transmission models to these data, accounting for number of contacts and baseline transmission probabilities, the pooled estimate suggests that the 20% most infectious cases have 3.1-fold (95% confidence interval: 2.2- to 4.2-fold) higher infectiousness than average cases, which is consistent with the observed heterogeneity in viral shedding. Household data can inform the estimation of transmission heterogeneity, which is important for epidemic management.

***For correspondence:**

timsang@connect.hku.hk (TKT);
bcowling@hku.hk (BJC)

[†]These authors contributed
equally to this work

Competing interest: See page 8

Funding: See page 8

Received: 11 August 2022

Preprinted: 30 August 2022

Accepted: 20 February 2023

Published: 07 March 2023

Reviewing Editor: Amy

Wesolowski, Johns Hopkins
Bloomberg School of Public
Health, United States

© Copyright Tsang *et al.* This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editor's evaluation

While it has been demonstrated that for SARS-CoV-2, a small fraction of individuals contributes to the majority of onward transmission, this heterogeneity is driven by multiple factors that span both biological and behavioral causes. By performing a solid meta-analysis of household transmission studies, the authors fit a household transmission model to the curated data to estimate variation in infectiousness which provides a valuable contribution to the existing knowledge base. By collating data from multiple studies, they are able to more fully investigate individual variability.

Introduction

Characterizing transmission is critical to control the spread of an emerging infectious disease. The reproductive number is the widely adopted measure of infectiousness. However, it only measures the average number of secondary cases infected by an infected person, not the heterogeneity in the number of transmissions. Variation of individual infectiousness is particularly highlighted by super-spreading events (SSEs), in which a minority of cases are responsible for a majority of transmission events. Such phenomena, illustrated by the '80/20 rule' (i.e., 20% of cases responsible for 80% transmission; *Adam et al., 2020; Sun et al., 2021*), have been observed in emerging infectious disease outbreaks (*Lloyd-Smith et al., 2005*), including severe acute respiratory syndrome (SARS) (*Shen et al., 2004*), Middle East respiratory syndrome (*Cauchemez et al., 2016; Cowling et al., 2015*), and most recently the COVID-19 pandemic (*Adam et al., 2020; Lau et al., 2020; Sun et al., 2021*);

Wong and Collins, 2020; Zhao et al., 2021). In these outbreaks, the proportion of cases attributed to 80% transmission, and the dispersion parameter that is estimated by fitting the negative binomial distribution to the number of secondary cases (Lloyd-Smith et al., 2005), are considered as measures of transmission heterogeneity.

However, the number of contacts per index cases is often not reported in SSE studies, and hence not incorporated in the analyses. In addition, SSE studies usually analyze clusters from different settings, in which the baseline transmission risk and density of exposure could be different (Thompson, 2021). Finally, studies of transmission heterogeneity that focus on SSEs described in the literature may suffer from publication bias, with larger clusters having higher probability of being observed and reported (Zhao et al., 2021). Therefore, the observed heterogeneity in the number of secondary cases could be the result of large number of contacts in SSE settings, or confounding from these factors (Althouse et al., 2020; Bagdasarian and Fisher, 2020), instead of variation in individual infectiousness.

Households are one of the most important settings for SARS-CoV-2 transmission, with 4- to 10-fold higher transmission risk than other places (Thompson, 2021). Hence, household transmission studies provide an ideal setting to quantify variations in individual infectiousness. In a household transmission study, an index case is identified, and their household contacts are followed up for 1–2 weeks, during which there is high transmission potential (Tsang et al., 2016). Therefore, the number of contacts is known while transmission risks and reporting biases can be controlled. We aim to characterize the variation of individual infectiousness by analyzing data from household transmission studies.

Results

We conducted a systematic review to gather information on the number of secondary cases with the number of household contacts for each household, in the form of number of households with X cases among households of size Y . In total, we identified 17 studies, comprising 13,098 index cases and 31,359 household contacts (Appendix 1—figure 1, Appendix 1—table 1; Bernardes-Souza et al., 2021; Carazo et al., 2022; Dattner et al., 2021; Han et al., 2022; Hart et al., 2022; Hsu et al., 2021; Hubiche et al., 2021; Koureas et al., 2021; Laxminarayan et al., 2020; Layan et al., 2022; Lyngse et al., 2022; Méndez-Echevarría et al., 2021; Posfay-Barbe et al., 2020; Reukers et al., 2021; Wilkinson et al., 2021). Most studies covered a period from January to November 2020, which was dominated by ancestral strains, except for Layan et al. (from December 2020 to April 2021) and Hsu et al. (from January 2020 to February 2021), which covered both ancestral strains and the alpha variant.

We then developed a statistical model to quantify the degree and the impact of variation of infectiousness of cases on transmission dynamics. The individual-based household transmission model describes the probability of infection of household contacts as depending on the time since infection in other infected persons in the household, so that infections from outside the household (community

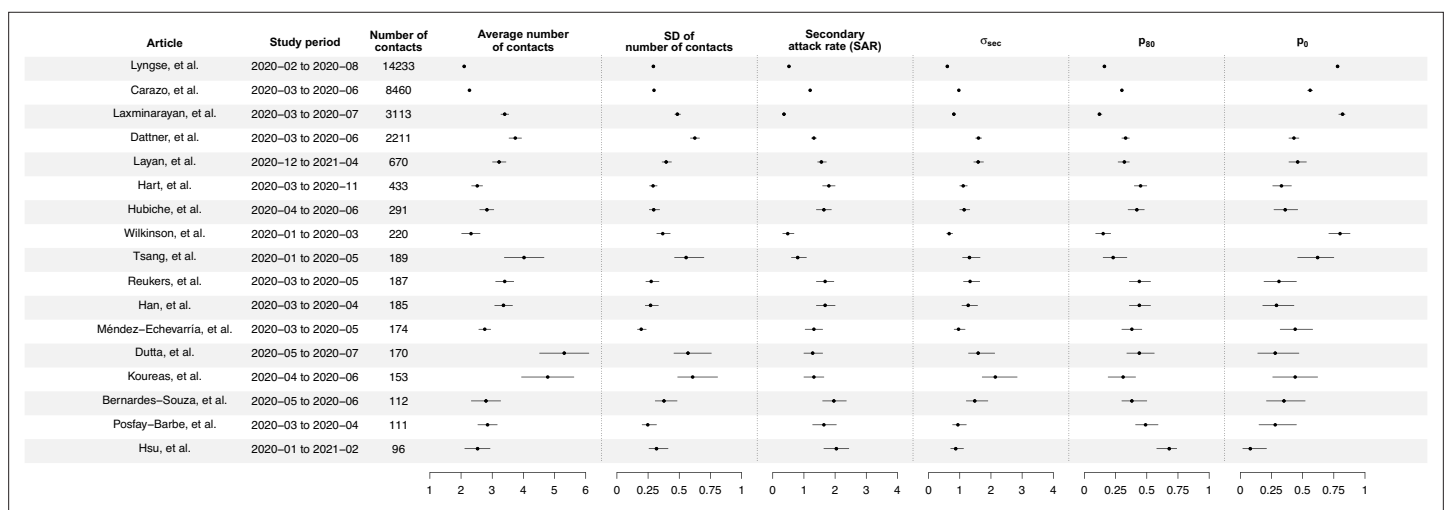


Figure 1. Summary of statistics for 17 identified studies. Figure shows the average number of contacts and standard deviation (SD) of number of contact, SD of number of secondary cases per index cases (σ_{sec}), and secondary attack rate (SAR) for 17 identified studies.

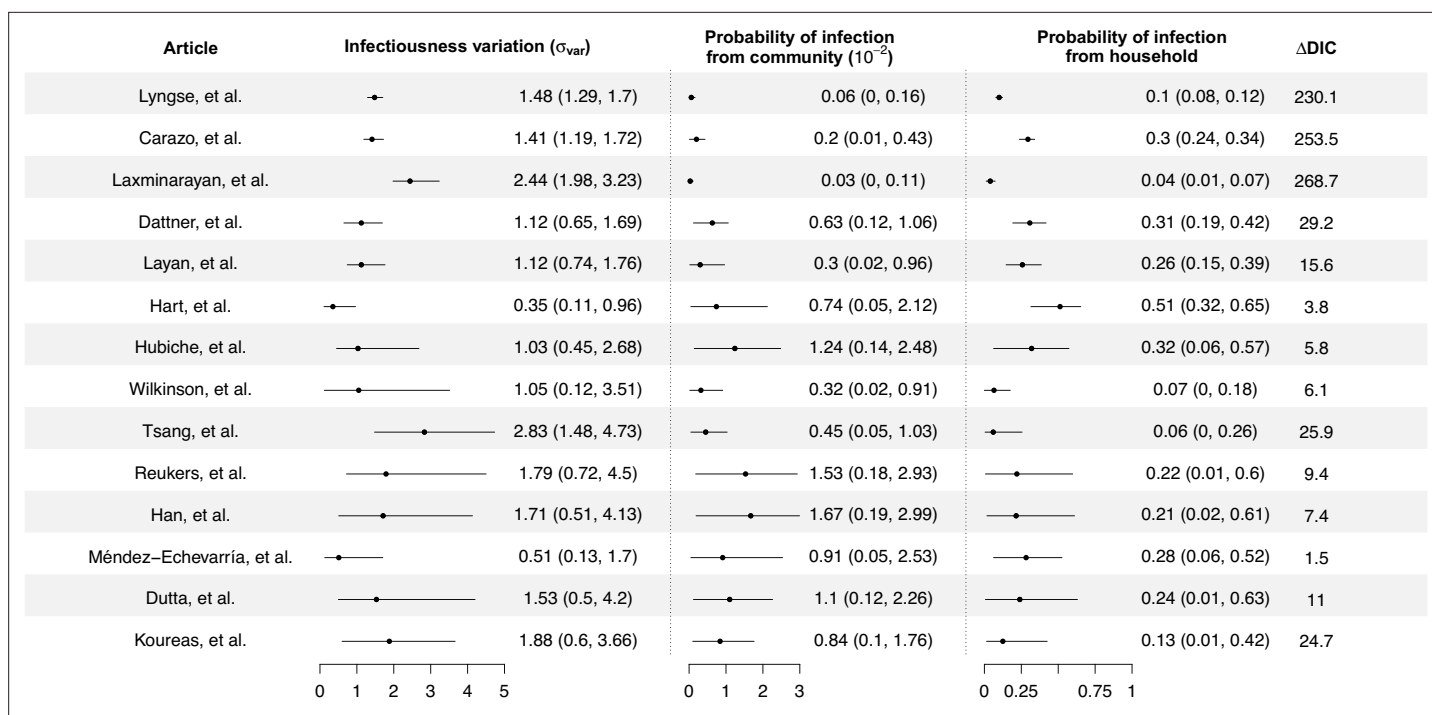


Figure 2. Modeling results of household transmission dynamics and infectiousness variation. Figure shows the estimates of infectiousness variation (σ_{var}), the estimated probability of infection from community and estimated probability of infection from households, and the reduction in deviance information criterion (DIC) compared with the model without infectiousness variation. Models are fitted separately to 14 household transmission studies.

infections), or infections via other household contacts rather than the index case (tertiary infections) are allowed (Cauchemez et al., 2009; Tsang et al., 2014; Tsang et al., 2021). Therefore, the model could estimate the per-contact hazard of infection, which implicitly controls for number of household contacts in households. We extend this model by adding a random effect (δ_i) on the individual infectiousness of cases. Here, the relative infectiousness of case i compared with case j is $\exp(\delta_i)/\exp(\delta_j)$. The parameter for variation in individual infectiousness (hereafter denoted as infectiousness variation, σ_{var}) is the standard deviation (SD) of the random effect characterizing individual infectiousness, so that δ_i follows a normal distribution with mean equal to 0 and SD equal to σ_{var} .

We separately fit the models to 14 studies with more than 150 contacts (Carazo et al., 2022; Dattner et al., 2021; Dutta et al., 2020; Han et al., 2022; Hart et al., 2022; Hubiche et al., 2021; Koureas et al., 2021; Laxminarayan et al., 2020; Layan et al., 2022; Lyngse et al., 2022; Méndez-Echevarría et al., 2021; Reukers et al., 2021; Tsang et al., 2022; Wilkinson et al., 2021; Figure 1, Appendix 1—table 2). For 12 studies out of 14, models with infectiousness variation perform substantially better (range of Δ DIC: 5.8–268) (Figure 2, Appendix 1—table 2). From these 12 studies, the estimated infectiousness variation (σ_{var}) ranged from 1.03 to 2.83. This suggests that, the 20% most infectious cases are 2.4- to 10-fold more infectious than the average case. Based on the two largest studies with 6782 and 3727 households (Carazo et al., 2022; Lyngse et al., 2022), the estimated infectiousness variation is 1.48 (95% credible interval [CrI]: 1.29, 1.7) and 1.41 (95% CrI: 1.19, 1.72), suggesting that, among all cases, the 20% most infectious are 3.5-fold (95% CrI: 3.0- to 4.2-fold) and 3.3-fold (95% CrI: 2.7- to 4.3-fold) more infectious than the average case. The estimated daily probability of infection from outside the household and estimated person-to-person transmission probability within households range from 0.0003 to 0.017, and from 0.06 to 0.51, respectively. The estimates of parameters for the relationship between number of contacts and transmission (larger value indicates stronger inverse association) range from 0.43 to 0.92, except for the study by Layan et al., 2022, where it is equal to 0.2. For all studies, the predicted final size distribution is consistent with the observed data and the model fit is judged adequate (Appendix 1—Tables 3–7, Appendix 1—figure 2). Simulation studies demonstrated that there is no important systematic bias, with 88–100% (depending on the parameter)

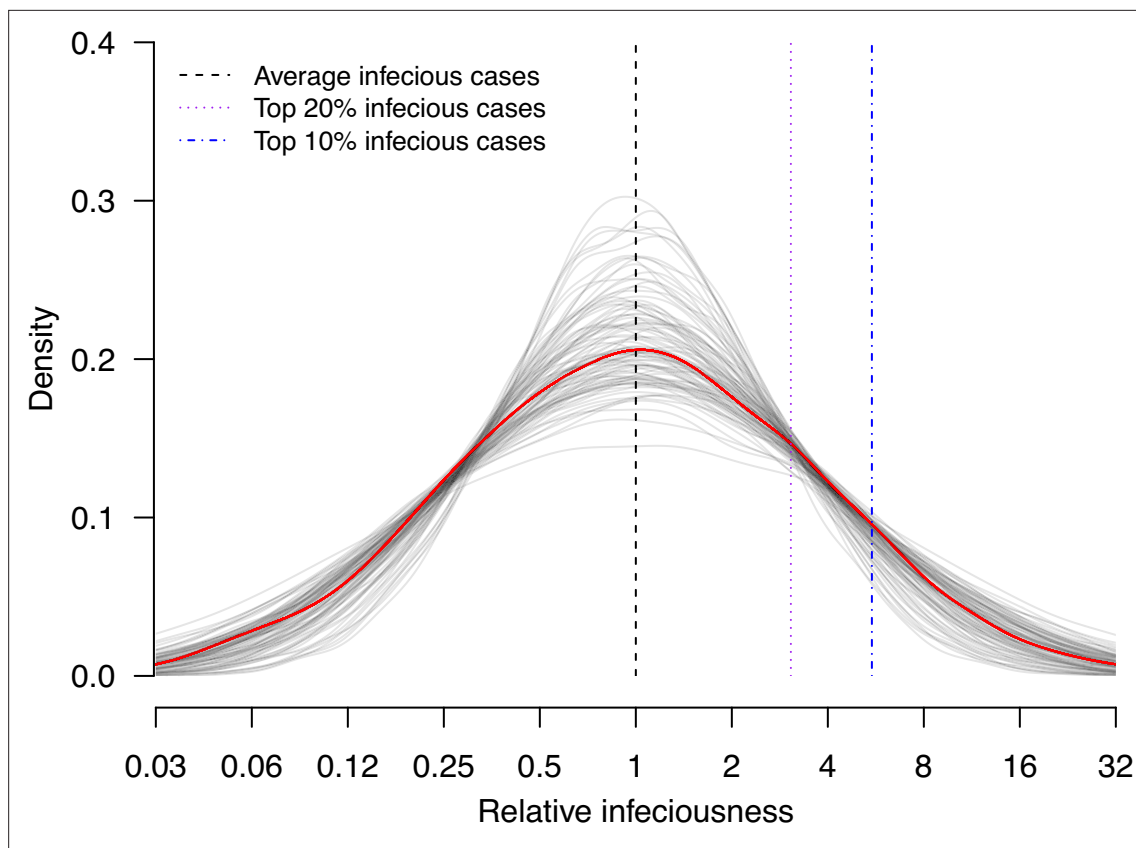


Figure 3. Estimate distribution of relative infectiousness based on the pooled estimate. Red line indicates the estimated distribution and the gray lines indicate the associated uncertainty. Black dashed line indicates average infectiousness (relative infectiousness equal to 1), while the purple and blue dashed lines indicate top 20% and 10% infectiousness, respectively.

of the 95% credible intervals covering the simulation value (**Appendix 1—table 8**). This suggests that the algorithm could estimate adequately the posterior distribution. We conduct a sensitivity analysis that assumes the individual infectiousness of cases follows a Gamma distribution, but the fit worsens substantially (**Appendix 1—table 9**).

We conduct random effects meta-analyses on estimates of individual infectiousness from the 14 identified studies. The pooled estimate of infectiousness variation is 1.33 (95% confidence interval [CI]: 0.95, 1.70), suggesting that the 20% most infectious cases are 3.1-fold (95% CI: 2.2- to 4.2-fold) more infectious than the average case (**Figure 3**). Based on this fitted distribution, we estimate that 5.9% (95% CI: 1.4%, 11.1%) and 14.9% (95% CI: 7.2%, 20.7%) of cases could be at least 8- and 4-fold more infectious than average cases, respectively.

We further explore if the secondary attack rate (SAR, the proportion of infected contacts), and the SD of the distribution of number of secondary cases (σ_{sec} (**Figure 4**; **Appendix 1—table 10**) may be correlated with the infectiousness variation. In meta-regression, we find the infectiousness variation is associated with SAR (**Appendix 1—table 11**). We estimate that doubling SAR are associated with 0.55 (95% CI: 0.21, 0.89) unit decrease in infectiousness variation, with R-squared equal to 67% respectively. In addition, we find that higher infectiousness variation is associated with only using PCR to ascertain secondary cases. Regarding other statistics, we find that σ_{sec} is positively associated with mean and SD of number of contacts. Other than these associations, we find no association between these statistics and implementation of lockdown, ascertainment method of index and secondary cases, and the circulating virus of SARS-CoV-2 in the study period.

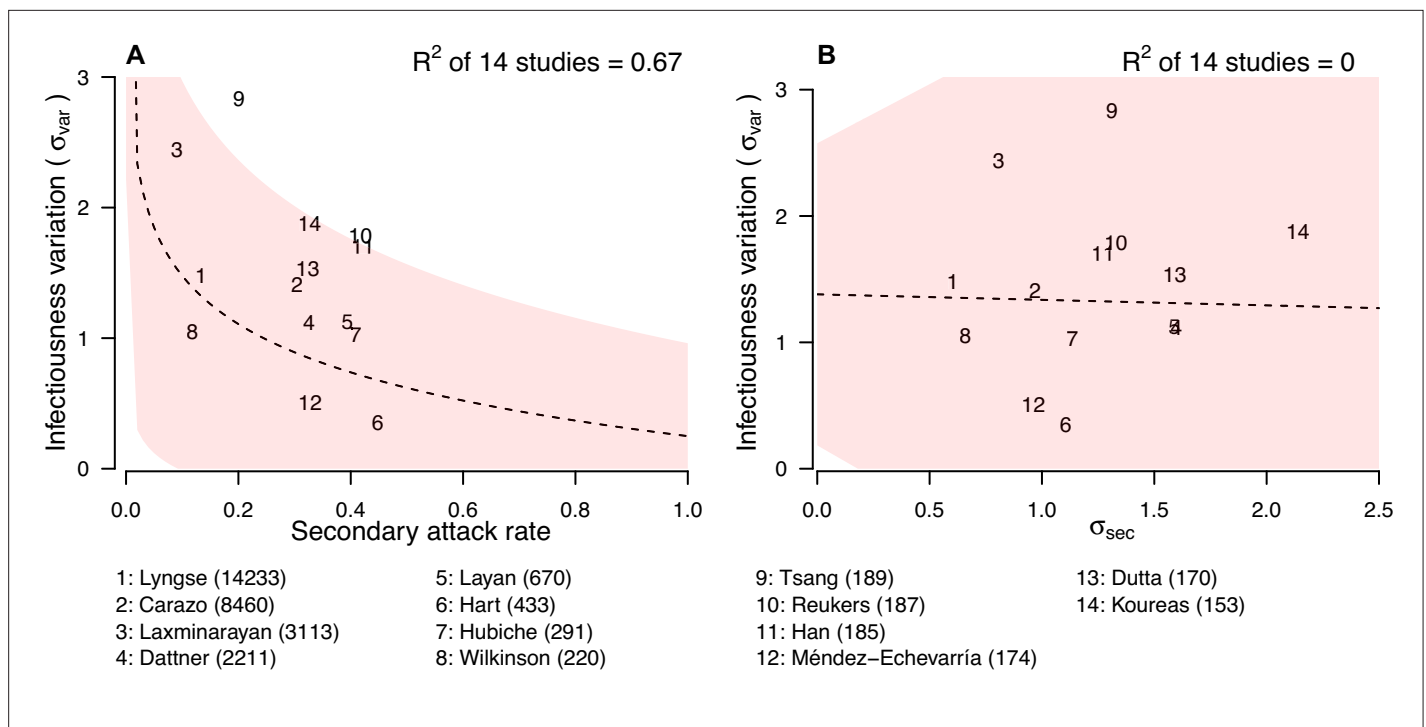


Figure 4. Relationship between infectiousness variation and statistic. In each panel, numbers represent the observed corresponding relationship for the identified studies. Panels A and B show the relationship between infectiousness variation (σ_{var}) and secondary attack rate (SAR) and standard deviation (SD) of number of secondary cases per index cases (σ_{sec}). In the bottom, numbers in bracket indicate the number of household contacts in corresponding studies.

Discussion

In this study, we characterize the impact of variation of individual infectiousness on heterogeneity of transmission of COVID-19 in households. We demonstrate that it can be estimated from household data using a modeling approach. The pooled estimate of infectiousness variation from 14 studies suggests that the 20% most infectious cases have 3.1-fold (95% CI: 2.2- to 4.2-fold) higher infectiousness compared with average index cases. This implies there is substantial variation in individual infectiousness of cases in households. Given that we focused our analysis on households with known number of contacts in studies conducted in the early outbreaks caused by ancestral strains, the estimated infectiousness variations are corrected for the variations caused by number of contacts and transmission risks in different settings, difference in pre-existing immunity among contacts (almost everyone is naïve and unvaccinated), and the difference in transmissibility among variants. Hence, this estimated infectiousness variation measures the variation caused by difference in individuals, which may be contributed by both biological factors and host behaviors, but not other potential confounding factors mentioned above.

Regarding host behaviors, multiple contact patterns, particularly by age, could contribute to the variations in infectiousness of cases. For example, mother-child contacts are usually more intense than father-child contacts, and adult cases are more capable of self-isolation within a household compared with children. Furthermore, contact pattern studies suggested that school-age children and young adults tended to mix with people of the same age (*Mossong et al., 2008; Mousa et al., 2021*). Also, the duration of contact could vary, which could also contribute to the variations in infectiousness of cases (*Toth et al., 2015*). It should be noted that contacts have different features, including their number, frequency, and duration that may contribute to variations in infectiousness. Previous studies also suggest that the relative importance of contact characteristics on transmission may be different among viruses (*De Cao et al., 2014*). Individual behaviors may also be influenced by control measures and recommendation from public health agencies. For example, lockdown and stay-at-home orders may increase the time spending at home. Mask-wearing when in contact with other household members, using separate bedrooms and bathrooms, avoiding having meals together may

reduce the transmission in households (**Ng et al., 2021**). Also, social disparity such as occupation may increase or reduce the risk of transmission in households, including the availability of personal protective equipment (PPE), or being healthcare worker (**Yang et al., 2021**). These factors may have impact of the transmission risk and hence SAR in households. In addition, heterogeneity in these different factors may have contributed infectiousness variation.

Biological factors may also contribute to such variations. For example, the SAR for index cases with fever and cough are 1.4- and 1.3-fold higher than index cases without fever and cough, respectively (**Madewell et al., 2021**). Viral shedding is used as a proxy measure of infectiousness, and consistently it also has substantial variations, as suggested by 11 systematic reviews (**Appendix 1—Tables 12–14**). Regarding the magnitude of viral shedding, studies report high variations of temporal viral shedding patterns among individuals (**He et al., 2020; Jones et al., 2021; Sun et al., 2022**). In addition, the duration of viral shedding can be highly heterogeneous, with pooled estimates of mean durations ranging from 11.1 to 30.3 days, and almost all reviews reporting high heterogeneity of estimates (**Cevik et al., 2021; Chen et al., 2021b; Díaz et al., 2022; Fontana et al., 2021; Li et al., 2020b; Okita et al., 2022; Qutub et al., 2022; Rahmani et al., 2022; Xu et al., 2020; Yan et al., 2021; Zhang et al., 2021**). Such heterogeneities still exist in subgroup analyses by age and severity (**Cevik et al., 2021; Fontana et al., 2021; Okita et al., 2022; Qutub et al., 2022; Rahmani et al., 2022; Xu et al., 2020; Yan et al., 2021**). The infectious period, proxied by duration of replicant competent virus isolation, is also heterogeneous (**Rahmani et al., 2022; Appendix 1—table 14**). One review also suggests that heterogeneity in viral shedding is an intrinsic virological factor facilitating higher dispersion parameter for SARS-CoV-2 if we compare it with the corresponding patterns in SARS-CoV-1 and pandemic influenza A(H1N1)pdm09 (**Chen et al., 2021a**).

The observed variation in individual infectiousness is consistent with past analyses of the dispersion parameter in a negative binomial distribution fitted to number of secondary cases per index case of COVID-19 (**Sun et al., 2021; Wong et al., 2015**). However, in these studies, the number of contacts was not considered. Therefore, the observed variations may not apply directly to households with limited number of contacts. It should be noted that the SD of the distribution of number of secondary cases (σ_{sec}) is only weakly correlated with infectiousness variation. This is because σ_{sec} is highly correlated with the mean and SD of the number of contacts, suggesting that it may depend on distribution of number of contacts, and hence may not be comparable among studies. Infectiousness variation is also correlated with the SAR. When SAR is higher, it is expected that more contacts are infected and therefore the observed number of secondary cases is less heterogeneous (**Adam et al., 2022**). We find higher infectiousness variation is associated with only using PCR to confirm secondary cases. One potential reason is that using other methods may lower the sensitivity of detecting infection, resulting in lower estimates of SAR and hence higher estimates of infectiousness variation. Further investigations are needed to explore roles of other potential factors affecting infectiousness variation, such as contact frequency among different regions (**Mousa et al., 2021**).

An important limitation of our study is that we do not have individual-level data. Therefore, we are unable to determine the impact of demographic factors like age and sex on infectiousness variation. Also, we cannot disentangle the host behaviors from biological factors. Previous analysis suggested no evidence of the impact of age on infectiousness of cases (**Lau et al., 2020; Sun et al., 2021**). In addition, we could not include factors affecting susceptibility to infection. Our estimates of infectiousness variation should be interpreted in light of these limitations: they capture heterogeneity in infectiousness due to demographic, host, and biological factors. However, in one study that included susceptibility component in the estimation of individual infectiousness, substantial heterogeneity remained with 20% of cases estimated to contribute to 80% of transmission (**Tsang et al., 2022**). Second, the recruitment methods among studies are different. This may affect the comparability of the results, although all index cases are laboratory-confirmed in all studies (**Tsang et al., 2021**). Third, we assumed that risks of infection from community for all households are the same, but there were different factors that may affecting this, including occupations, such as healthcare workers, social economic status that related to assess to PPEs (**Yang et al., 2021**). Finally, most of our identified studies were conducted in the period of circulation of ancestral strains, and therefore the identified infectiousness variation may not be directly applicable to other variants.

In conclusion, we developed a modeling approach to estimate variation in individual infectiousness from household data. Result indicates that there is substantial variation in individual infectiousness, which is important for epidemic management.

Materials and methods

Study design

The aim of this study was to develop a statistical model to quantify the variation of individual infectiousness in households, based on publicly available information. An index case was defined as the first detected case in a household, while secondary cases were defined as the identified infected household contacts of the index case. We conducted a systematic review to collect household studies with at least 30 households, reporting the number of secondary cases with number of household contacts for each household for COVID-19, in the form of number of households with X cases among households of size Y . For each study, we also extracted the study period, the coverage of tests of household contacts, the case ascertainment methods, the circulating virus of SARS-CoV-2, and the public health and social measures in the study period. This information was used as an input of modeling analyses in this study. Details of systematic review could be found in Appendix 1.

Estimation of variation of individual infectiousness in households

To determine if there are variations of individual infectiousness of cases, we used an individual-based household transmission model (; [Tsang et al., 2014](#); [Tsang et al., 2021](#)). The model described the probability of infection of household contacts as depending on the time since infection in other infected people in the household, while infections from outside the household (community infections), or infections via other household contacts rather than the index case (tertiary infections) are allowed. We extended the model by adding a random effect parameter (δ_i) on the individual infectiousness of each case. In the model, the hazard of infection of individual j at time t from an infected household member i , with infection time t_i in household k , was

$$\lambda_{i \rightarrow j}(t) = \frac{\lambda_i}{X_k^\beta} * \exp(\delta_i) * f(t - t_i)$$

where λ_i was the baseline hazard, δ_i followed a normal distribution with mean 0 and SD σ_{var} , which quantified the variation of individual infectiousness (hereafter denoted as infectiousness variation). The relative infectiousness of case i compared with case j was $\exp(\delta_i)/\exp(\delta_j)$.

X_k was the number of household contacts. β was the parameter describing the relationship between number of household contacts and transmission rate. It ranged from 0 to 1, with 0 indicating that the transmission rate was independent of number of household contacts while 1 indicated that the transmission rate was inversely proportional to number of household contacts (i.e. dilution effect of the contact time per contact which was lower when the number of household contacts is higher). $f(\cdot)$ was the infectiousness profile since infection generated from the assumed incubation period (mean equal to 5 days) and infectious period (mean equal to 13) ([He et al., 2020](#); [Jing et al., 2020](#); [Li et al., 2020a](#); [Appendix 1—table 15](#)). Extracted distribution were summarized in [Appendix 1—table 15](#), and the shape of the assumed function was plotted on [Appendix 1—figure 3](#).

Since the data were extracted from publication, the infection time of all cases was unavailable for all studies. Also, the individual infectiousness parameters δ_i for cases were augmented variables. Therefore, we conducted our inference under a Bayesian framework using data augmentation Markov chain Monte Carlo (MCMC) algorithm to joint estimate the model parameters, the missing infection time, and augmented variables using metropolis-hasting algorithm. We separately fitted this model to identified studies. We assessed the model adequacy by comparing the observed and expected number of infections in households by household size. We compared the model with or without the random effect for variation of individual infectiousness by using deviance information criterion (DIC) ([Spiegelhalter et al., 2002](#)). Smaller DIC indicated a better model fit. DIC difference >5 was considered as substantial improvement ([Spiegelhalter et al., 2000](#)). Details of the model and inference could be found in Appendix 1. We conducted a sensitivity analysis that used a Gamma distribution, instead of exponential of normal distribution (Appendix 1).

Meta-analysis and meta-regression

We conducted random effects meta-analyses on identified studies to obtain pooled estimates of individual infectiousness, using the inverse variance method and restricted maximum likelihood estimator for heterogeneity (*Hedges and Vevea, 1998; Langan et al., 2019; Thompson and Sharp, 1999; Veroniki et al., 2019*). Cochran Q test and the I^2 statistic were used to identify and quantify heterogeneity among included studies (*Cochran, 1954; Higgins et al., 2003*). An I^2 value more than 75% indicates high heterogeneity (*Higgins et al., 2003*). We conducted a meta-regression analysis to explore the association between infectiousness variation (σ_{var} , SD) of the distribution of number of secondary cases (σ_{sec} , and SAR), and further including the following factors: the mean and SD of number of household contacts, implementation of lockdown, ascertainment method of index and secondary cases, only ancestral strains are circulating in study period, and all household contacts were tested.

Data availability

All data in this study are publicly available since they are extracted from published articles. Summarized data for analysis could be downloaded from https://github.com/timktsang/covid19_transmission_heterogeneity, (copy archived at [swh:1:rev:634390fa66a4bfb998da691e7cd81cf45e38c6db](https://www.swh.io/rev/634390fa66a4bfb998da691e7cd81cf45e38c6db); *Tsang, 2022*).

Acknowledgements

The authors thank Maylis Layan for helpful discussion and Jiayi Zhou for technical assistance. 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; BJC) and the Collaborative Research Fund (Project No. C7123-20G; BJC) of the Research Grants Council of the Hong Kong SAR Government. BJC is supported by the AIR@innoHK program of the Innovation and Technology Commission of the Hong Kong SAR Government. SC acknowledges financial support from the Investissement d'Avenir program, the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases program (grant ANR-10-LABX-62-IBEID), the EMERGEN project (ANRS0151), the INCEPTION project (PIA/ANR-16-CONV-0005), the European Union's Horizon 2020 research and innovation program under grant 101003589 (RECOVER) and 874735 (VEO), AXA and Groupama. TKT acknowledges the Seed Fund for Basic Research (202111159118) from the University of Hong Kong.

Additional information

Competing interests

Benjamin John Cowling: consults for AstraZeneca, Fosun Pharma, GlaxoSmithKline, Moderna, Pfizer, Roche and Sanofi Pasteur. The authors report no other potential conflicts of interest. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
Food and Health Bureau	COVID190118	Benjamin John Cowling
Research Grants Council, University Grants Committee	C7123-20G	Benjamin John Cowling
Innovation and Technology Commission of Hong Kong Special Administrative Government	AIR@innoHK	Benjamin John Cowling
Investissement d'Avenir	ANR-10-LABX-62-IBEID	Simon Cauchemez
EMERGEN	ANRS0151	Simon Cauchemez
INCEPTION	PIA/ANR-16-CONV-0005	Simon Cauchemez

Funder	Grant reference number	Author
AXA Research Fund	101003589 (RECOVER)	Simon Cauchemez
Groupama	874735 (VEO)	Simon Cauchemez
The University of Hong Kong	202111159118	Tim K Tsang

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Tim K Tsang, Conceptualization, Data curation, Software, Formal analysis, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing; Xiaotong Huang, Data curation, Investigation, Visualization; Can Wang, Data curation, Investigation; Sijie Chen, Investigation, Visualization; Bingyi Yang, Writing - review and editing; Simon Cauchemez, Benjamin John Cowling, Conceptualization, Supervision, Funding acquisition, Methodology, Writing - review and editing

Author ORCIDs

Tim K Tsang  <http://orcid.org/0000-0001-5037-6776>

Simon Cauchemez  <http://orcid.org/0000-0001-9186-4549>

Benjamin John Cowling  <http://orcid.org/0000-0002-6297-7154>

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.82611.sa1>

Author response <https://doi.org/10.7554/eLife.82611.sa2>

Additional files

Supplementary files

- MDAR checklist

Data availability

The data and computer code (in R language) for conducting the data analysis can be downloaded from https://github.com/timktsang/covid19_transmission_heterogeneity (copy archived at [swh:1:rev:634390fa66a4bfb998da691e7cd81cf45e38c6db](https://swh.io/rev/634390fa66a4bfb998da691e7cd81cf45e38c6db)).

References

- Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, Leung GM, Cowling BJ. 2020. Clustering and superspreading potential of SARS-cov-2 infections in hong kong. *Nature Medicine* **26**:1714–1719. DOI: <https://doi.org/10.1038/s41591-020-1092-0>, PMID: 32943787
- Adam D, Gostic K, Tsang T, Wu P, Lim WW, Yeung A, Wong J, Lau E, Du Z, Chen D, Ho LM, Martín-Sánchez M, Cauchemez S, Cobey S, Leung G, Cowling B. 2022. Time-Varying Transmission Heterogeneity of SARS and COVID-19 in Hong Kong. *Research Square*. DOI: <https://doi.org/10.21203/rs.3.rs-1407962/v1>
- Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, Hu H. 2020. Superspreading events in the transmission dynamics of SARS-cov-2: opportunities for interventions and control. *PLOS Biology* **18**:e3000897. DOI: <https://doi.org/10.1371/journal.pbio.3000897>, PMID: 33180773
- Bagdasarjan N, Fisher D. 2020. Heterogenous COVID-19 transmission dynamics within singapore: a clearer picture of future national responses. *BMC Medicine* **18**:164. DOI: <https://doi.org/10.1186/s12916-020-01625-7>, PMID: 32493311
- Bernardes-Souza B, Júnior SRC, Santos CA, Neto R, Bottega FDC, Godoy DC, Freitas BL, Silva DLG, Brinker TJ, Nascimento RA, Tupinambás U, Reis AB, Coura-Vital W. 2021. Logistics workers are a key factor for SARS-cov-2 spread in brazilian small towns: case-control study. *JMIR Public Health and Surveillance* **7**:e30406. DOI: <https://doi.org/10.2196/30406>, PMID: 34388105
- Carazo S, Laliberté D, Villeneuve J, Martin R, Deshaies P, Denis G, Adib G, Tissot F, Dionne M, De Serres G. 2022. Characterization and evolution of infection control practices among severe acute respiratory coronavirus virus 2 (SARS-cov-2)-infected healthcare workers in acute-care hospitals and long-term care facilities in québec, canada, spring 2020. *Infection Control and Hospital Epidemiology* **43**:481–489. DOI: <https://doi.org/10.1017/ice.2021.160>, PMID: 33853702

- Cauchemez S**, Donnelly CA, Reed C, Ghani AC, Fraser C, Kent CK, Finelli L, Ferguson NM. 2009. Household transmission of 2009 pandemic influenza A (H1N1) virus in the united states. *The New England Journal of Medicine* **361**:2619–2627. DOI: <https://doi.org/10.1056/NEJMoa0905498>, PMID: 20042753
- Cauchemez S**, Ferguson NM, Fox A, Mai LQ, Thanh LT, Thai PQ, Thoang DD, Duong TN, Minh Hoa LN, Tran Hien N, Horby P. 2014. Determinants of influenza transmission in south east asia: insights from a household cohort study in vietnam. *PLOS Pathogens* **10**:e1004310. DOI: <https://doi.org/10.1371/journal.ppat.1004310>, PMID: 25144780
- Cauchemez S**, Nouvellet P, Cori A, Jombart T, Garske T, Clapham H, Moore S, Mills HL, Salje H, Collins C, Rodriguez-Barraquer I, Riley S, Truelove S, Algarni H, Alhakeem R, AlHarbi K, Turkistani A, Aguas RJ, Cummings DAT, Van Kerkhove MD, et al. 2016. Unraveling the drivers of MERS-cov transmission. *PNAS* **113**:9081–9086. DOI: <https://doi.org/10.1073/pnas.1519235113>, PMID: 27457935
- Celex G**, Forbes F, Robert CP, Titterton DM. 2006. Deviance information criteria for missing data models. *Bayesian Analysis* **1**:651–673. DOI: <https://doi.org/10.1214/06-BA122>
- Cevik M**, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. 2021. SARS-cov-2, SARS-cov, and MERS-cov viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet. Microbe* **2**:e13–e22. DOI: [https://doi.org/10.1016/S2666-5247\(20\)30172-5](https://doi.org/10.1016/S2666-5247(20)30172-5), PMID: 33521734
- Chen PZ**, Bobrovitz N, Premji Z, Koopmans M, Fisman DN, Gu FX. 2021a. Heterogeneity in transmissibility and shedding SARS-cov-2 via droplets and aerosols. *eLife* **10**:e65774. DOI: <https://doi.org/10.7554/eLife.65774>, PMID: 33861198
- Chen C**, Zhu C, Yan D, Liu H, Li D, Zhou Y, Fu X, Wu J, Ding C, Tian G, Lan L, Liu X, Huang C, Hecht R, Li L, Yang S. 2021b. The epidemiological and radiographical characteristics of asymptomatic infections with the novel coronavirus (COVID-19): a systematic review and meta-analysis. *International Journal of Infectious Diseases* **104**:458–464. DOI: <https://doi.org/10.1016/j.ijid.2021.01.017>, PMID: 33444755
- Cochran WG**. 1954. The combination of estimates from different experiments. *Biometrics* **10**:101. DOI: <https://doi.org/10.2307/3001666>
- Cowling BJ**, Park M, Fang VJ, Wu P, Leung GM, Wu JT. 2015. Preliminary epidemiological assessment of MERS-cov outbreak in south korea, may to june 2015. *Euro Surveillance* **20**:7–13. DOI: <https://doi.org/10.2807/1560-7917.es2015.20.25.21163>, PMID: 26132767
- Dattner I**, Goldberg Y, Katriel G, Yaari R, Gal N, Miron Y, Ziv A, Sheffer R, Hamo Y, Huppert A. 2021. The role of children in the spread of COVID-19: using household data from bnei BRAK, israel, to estimate the relative susceptibility and infectivity of children. *PLOS Computational Biology* **17**:e1008559. DOI: <https://doi.org/10.1371/journal.pcbi.1008559>, PMID: 33571188
- De Cao E**, Zagheni E, Manfredi P, Melegaro A. 2014. The relative importance of frequency of contacts and duration of exposure for the spread of directly transmitted infections. *Biostatistics* **15**:470–483. DOI: <https://doi.org/10.1093/biostatistics/kxu008>, PMID: 24705143
- Díaz LA**, García-Salum T, Fuentes-López E, Reyes D, Ortiz J, Chahuan J, Levican J, Almonacid LI, Valenzuela GH, Serrano E, Budnik S, Gandara V, Gallardo A, Seydewitz MF, Ferrés M, Cofré C, Álvarez M, Pavez C, Candia R, Monroy H, et al. 2022. High prevalence of SARS-cov-2 detection and prolonged viral shedding in stools: a systematic review and cohort study. *Gastroenterologia y Hepatologia* **45**:593–604. DOI: <https://doi.org/10.1016/j.gastrohep.2021.12.009>, PMID: 35077722
- Dutta S**, Kaur RJ, Bhardwaj P, Charan J, Bist SKS, Detha MD, Kanchan T, Sharma P, Misra S. 2020. Household transmission of COVID-19: a cross-sectional study. *Infection and Drug Resistance* **13**:4637–4642. DOI: <https://doi.org/10.2147/IDR.S285446>, PMID: 33380813
- Fontana LM**, Villamagna AH, Sikka MK, McGregor JC. 2021. Understanding viral shedding of severe acute respiratory coronavirus virus 2 (SARS-cov-2): review of current literature. *Infection Control and Hospital Epidemiology* **42**:659–668. DOI: <https://doi.org/10.1017/ice.2020.1273>, PMID: 33077007
- Han WGH**, Swart A, Bonačić Marinović A, Eggink D, Reimerink J, Wijsman LA, van der Veer B, van den Brink S, van den Brandt A-M, van Tol S, Godeke G-J, Brouwer F, Hoogerwerf M, Reukers DFM, Rots N, Reusken C, Meijer A, Dutch F. F. X.-COVID-19 Research Group. 2022. SARS-cov-2 RNA and antibody dynamics in a dutch household study with dense sampling frame. *Scientific Reports* **12**:7937. DOI: <https://doi.org/10.1038/s41598-022-11480-6>, PMID: 35562380
- Hart WS**, Abbott S, Endo A, Hellewell J, Miller E, Andrews N, Maini PK, Funk S, Thompson RN. 2022. Inference of the SARS-cov-2 generation time using UK household data. *eLife* **11**:e70767. DOI: <https://doi.org/10.7554/eLife.70767>, PMID: 35138250
- He X**, Lau EHY, Wu P, Deng X, Wang J, Hao X, Lau YC, Wong JY, Guan Y, Tan X, Mo X, Chen Y, Liao B, Chen W, Hu F, Zhang Q, Zhong M, Wu Y, Zhao L, Zhang F, et al. 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* **26**:672–675. DOI: <https://doi.org/10.1038/s41591-020-0869-5>, PMID: 32296168
- Hedges LV**, Vevea JL. 1998. Fixed- and random-effects models in meta-analysis. *Psychological Methods* **3**:486–504. DOI: <https://doi.org/10.1037/1082-989X.3.4.486>
- Higgins JPT**, Thompson SG, Deeks JJ, Altman DG. 2003. Measuring inconsistency in meta-analyses. *BMJ* **327**:557–560. DOI: <https://doi.org/10.1136/bmj.327.7414.557>, PMID: 12958120
- Hsu CY**, Wang JT, Huang KC, Fan ACH, Yeh YP, Chen SLS. 2021. Household transmission but without the community-acquired outbreak of COVID-19 in taiwan. *Journal of the Formosan Medical Association* **120 Suppl 1**:S38–S45. DOI: <https://doi.org/10.1016/j.jfma.2021.04.021>, PMID: 33994234
- Hubiche T**, Phan A, Leducq S, Rapp J, Fertitta L, Aubert H, Barbarot S, Chiaverini C, Giraudeau B, Lasek A, Mallet S, Labarelle A, Piram M, McCuaig C, Martin L, Monitor L, Nicol I, Bissuel M, Bellissen A, Jullien D, et al. 2021. Acute

- acral eruptions in children during the COVID-19 pandemic: characteristics of 103 children and their family clusters. *Annales de Dermatologie et de Venereologie* **148**:94–100. DOI: <https://doi.org/10.1016/j.annder.2020.11.005>, PMID: 33551211
- Jing QL, Liu MJ, Zhang ZB, Fang LQ, Yuan J, Zhang AR, Dean NE, Luo L, Ma MM, Longini I, Kenah E, Lu Y, Ma Y, Jalali N, Yang ZC, Yang Y. 2020. Household secondary attack rate of COVID-19 and associated determinants in guangzhou, china: a retrospective cohort study. *The Lancet. Infectious Diseases* **20**:1141–1150. DOI: [https://doi.org/10.1016/S1473-3099\(20\)30471-0](https://doi.org/10.1016/S1473-3099(20)30471-0), PMID: 32562601
- Jones TC, Biele G, Mühlemann B, Veith T, Schneider J, Beheim-Schwarzbach J, Bleicker T, Tesch J, Schmidt ML, Sander LE, Kurth F, Menzel P, Schwarzer R, Zuchowski M, Hofmann J, Krumbholz A, Stein A, Edelmann A, Corman VM, Drosten C. 2021. Estimating infectiousness throughout SARS-cov-2 infection course. *Science* **373**:eabi5273. DOI: <https://doi.org/10.1126/science.abi5273>, PMID: 34035154
- Koureas M, Speletas M, Bogogiannidou Z, Babalis D, Pinakas V, Pinaka O, Komnos A, Tsoutsas S, Papadamou G, Kyritsi MA, Vontas A, Nakoulas V, Sapounas S, Kanellopoulos N, Kalompatsios D, Papadoulis V, Dadouli K, Soteriades S, Mina P, Mouchtouri VA, et al. 2021. Transmission dynamics of SARS-cov-2 during an outbreak in a roma community in thessaly, greece-control measures and lessons learned. *International Journal of Environmental Research and Public Health* **18**:2878. DOI: <https://doi.org/10.3390/ijerph18062878>, PMID: 33799791
- Langan D, Higgins JPT, Jackson D, Bowden J, Veroniki AA, Kontopantelis E, Viechtbauer W, Simmonds M. 2019. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods* **10**:83–98. DOI: <https://doi.org/10.1002/jrsm.1316>, PMID: 30067315
- Lau MSY, Grenfell B, Thomas M, Bryan M, Nelson K, Lopman B. 2020. Characterizing superspreading events and age-specific infectiousness of SARS-cov-2 transmission in georgia, USA. *PNAS* **117**:22430–22435. DOI: <https://doi.org/10.1073/pnas.2011802117>, PMID: 32820074
- Laxminarayan R, Wahl B, Dudala SR, Gopal K, Mohan B C, Neelima S, Jawahar Reddy KS, Radhakrishnan J, Lewnard JA. 2020. Epidemiology and transmission dynamics of COVID-19 in two indian states. *Science* **370**:691–697. DOI: <https://doi.org/10.1126/science.abd7672>, PMID: 33154136
- Layan M, Gilboa M, Gonen T, Goldenfeld M, Meltzer L, Andronico A, Hozé N, Cauchemez S, Regev-Yochay G. 2022. Impact of bnt162b2 vaccination and isolation on SARS-cov-2 transmission in israeli households: an observational study. *American Journal of Epidemiology* **191**:1224–1234. DOI: <https://doi.org/10.1093/aje/kwac042>, PMID: 35238329
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, et al. 2020a. Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. *The New England Journal of Medicine* **382**:1199–1207. DOI: <https://doi.org/10.1056/NEJMoa2001316>, PMID: 31995857
- Li B, Zhang S, Zhang R, Chen X, Wang Y, Zhu C. 2020b. Epidemiological and clinical characteristics of COVID-19 in children: A systematic review and meta-analysis. *Frontiers in Pediatrics* **8**:591132. DOI: <https://doi.org/10.3389/fped.2020.591132>, PMID: 33224909
- Liu JS. 2001. *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag.
- Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**:355–359. DOI: <https://doi.org/10.1038/nature04153>, PMID: 16292310
- Lyngse FP, Kirkeby C, Halasa T, Andreasen V, Skov RL, Møller FT, Krause TG, Mølbak K. 2022. Nationwide study on SARS-cov-2 transmission within households from lockdown to reopening, denmark, 27 february 2020 to 1 august 2020. *Euro Surveillance* **27**:2001800. DOI: <https://doi.org/10.2807/1560-7917.ES.2022.27.6.2001800>, PMID: 35144726
- Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE. 2021. Factors associated with household transmission of SARS-cov-2: an updated systematic review and meta-analysis. *JAMA Network Open* **4**:e2122240. DOI: <https://doi.org/10.1001/jamanetworkopen.2021.22240>, PMID: 34448865
- Méndez-Echevarría A, Sainz T, de Felipe B, Alcolea S, Olbrich P, Goycochea-Valdivia WA, Escosa-García L, Cobo L, Calvo C, Neth O. 2021. High rates of SARS-cov-2 family transmission in children of healthcare workers during the first pandemic wave in madrid, spain: serologic study. *The Pediatric Infectious Disease Journal* **40**:e185–e188. DOI: <https://doi.org/10.1097/INF.0000000000003088>, PMID: 33591078
- Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* **339**:b2535. DOI: <https://doi.org/10.1136/bmj.b2535>, PMID: 19622551
- Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ. 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine* **5**:e74. DOI: <https://doi.org/10.1371/journal.pmed.0050074>, PMID: 18366252
- Mousa A, Winskill P, Watson OJ, Ratmann O, Monod M, Ajelli M, Diallo A, Dodd PJ, Grijalva CG, Kiti MC, Krishnan A, Kumar R, Kumar S, Kwok KO, Lanata CF, de Waroux OLP, Leung K, Mahikul W, Melegaro A, Morrow CD, et al. 2021. Social contact patterns and implications for infectious disease transmission - a systematic review and meta-analysis of contact surveys. *eLife* **10**:e70294. DOI: <https://doi.org/10.7554/eLife.70294>, PMID: 34821551
- Ng OT, Marimuthu K, Koh V, Pang J, Linn KZ, Sun J, De Wang L, Chia WN, Tiu C, Chan M, Ling LM, Vasoo S, Abdad MY, Chia PY, Lee TH, Lin RJ, Sadarangani SP, Chen MI-C, Said Z, Kurupatham L, et al. 2021. SARS-cov-2 seroprevalence and transmission risk factors among high-risk close contacts: a retrospective cohort study. *The Lancet. Infectious Diseases* **21**:333–343. DOI: [https://doi.org/10.1016/S1473-3099\(20\)30833-1](https://doi.org/10.1016/S1473-3099(20)30833-1), PMID: 33152271

- Okita Y, Morita T, Kumanogoh A.** 2022. Duration of SARS-cov-2 RNA positivity from various specimens and clinical characteristics in patients with COVID-19: a systematic review and meta-analysis. *Inflammation and Regeneration* **42**:16. DOI: <https://doi.org/10.1186/s41232-022-00205-x>, PMID: 35642011
- Posfay-Barbe KM, Wagner N, Gauthey M, Moussaoui D, Loevy N, Diana A, L'Huillier AG.** 2020. COVID-19 in children and the dynamics of infection in families. *Pediatrics* **146**:e20201576. DOI: <https://doi.org/10.1542/peds.2020-1576>, PMID: 32457213
- Outub M, Aldabbagh Y, Mehdawi F, Alraddadi A, Alhomsi M, Alnahdi A, Fakeeh M, Maghrabi A, Alwagdani M, Bahabri N.** 2022. Duration of viable SARS-cov-2 shedding from respiratory tract in different human hosts and its impact on isolation discontinuation policies revision; a narrative review. *Clinical Infection in Practice* **13**:100140. DOI: <https://doi.org/10.1016/j.clinpr.2022.100140>, PMID: 35190799
- Rahmani A, Dini G, Leso V, Montecucco A, Kuszniir Vitturi B, Iavicoli I, Durando P.** 2022. Duration of SARS-cov-2 shedding and infectivity in the working age population: a systematic review and meta-analysis. *La Medicina Del Lavoro* **113**:e2022014. DOI: <https://doi.org/10.23749/ml.v113i2.12724>, PMID: 35481581
- Reukers DFM, van Boven M, Meijer A, Rots N, Reusken C, Roof I, van Gageldonk-Lafeber AB, van der Hoek W, van den Hof S.** 2021. High infection attack rates of SARS-cov-2 in dutch households revealed by dense sampling. *Infectious Diseases* **74**:52–58. DOI: <https://doi.org/10.1101/2021.01.26.21250512>
- Shen Z, Ning F, Zhou W, He X, Lin C, Chin DP, Zhu Z, Schuchat A.** 2004. Superspreading SARS events, Beijing, 2003. *Emerging Infectious Diseases* **10**:256–260. DOI: <https://doi.org/10.3201/eid1002.030732>, PMID: 15030693
- Spiegelhalter DJ, Thomas A, Best NG.** 2000. *WinBUGS Version 1.3 User Manual*. Cambridge: Medical Research Council Biostatistics Unit.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A.** 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society* **64**:583–639. DOI: <https://doi.org/10.1111/1467-9868.00353>
- Sun K, Wang W, Gao L, Wang Y, Luo K, Ren L, Zhan Z, Chen X, Zhao S, Huang Y, Sun Q, Liu Z, Litvinova M, Vespignani A, Ajelli M, Viboud C, Yu H.** 2021. Transmission heterogeneities, kinetics, and controllability of SARS-cov-2. *Science* **371**:eabe2424. DOI: <https://doi.org/10.1126/science.abe2424>, PMID: 33234698
- Sun K, Tempia S, Kleynhans J, von Gottberg A, McMorrow ML, Wolter N, Bhiman JN, Moyes J, du Plessis M, Carrim M, Buys A, Martinson NA, Kahn K, Tollman S, Lebina L, Wafawanaka F, du Toit JD, Gómez-Olivé FX, Mkhencele T, Viboud C, et al.** 2022. SARS-cov-2 transmission, persistence of immunity, and estimates of omicron's impact in south african population cohorts. *Science Translational Medicine* **14**:eabo7081. DOI: <https://doi.org/10.1126/scitranslmed.abo7081>, PMID: 35638937
- Thompson SG, Sharp SJ.** 1999. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* **18**:2693–2708. DOI: [https://doi.org/10.1002/\(sici\)1097-0258\(19991030\)18:20<2693::aid-sim235>3.0.co;2-v](https://doi.org/10.1002/(sici)1097-0258(19991030)18:20<2693::aid-sim235>3.0.co;2-v), PMID: 10521860
- Thompson HA.** 2021. SARS-cov-2 setting-specific transmission rates: a systematic review and meta-analysis. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* **73**:e754–e764. DOI: <https://doi.org/10.1093/cid/ciab100>
- Toth DJA, Leecaster M, Pettey WBP, Gundlapalli AV, Gao H, Rainey JJ, Uzicanin A, Samore MH.** 2015. The role of heterogeneity in contact timing and duration in network models of influenza spread in schools. *Journal of the Royal Society, Interface* **12**:20150279. DOI: <https://doi.org/10.1098/rsif.2015.0279>, PMID: 26063821
- Tsang TK, Cauchemez S, Perera RAPM, Freeman G, Fang VJ, Ip DKM, Leung GM, Malik Peiris JS, Cowling BJ.** 2014. Association between antibody titers and protection against influenza virus infection within households. *The Journal of Infectious Diseases* **210**:684–692. DOI: <https://doi.org/10.1093/infdis/jiu186>, PMID: 24676208
- Tsang TK, Lau LLH, Cauchemez S, Cowling BJ.** 2016. Household transmission of influenza virus. *Trends in Microbiology* **24**:123–133. DOI: <https://doi.org/10.1016/j.tim.2015.10.012>, PMID: 26612500
- Tsang TK, Wang C, Yang B, Cauchemez S, Cowling BJ.** 2021. Using secondary cases to characterize the severity of an emerging or re-emerging infection. *Nature Communications* **12**:6372. DOI: <https://doi.org/10.1038/s41467-021-26709-7>, PMID: 34737277
- Tsang T.** 2022. Covid19_transmission_heterogeneity. swh:1:rev:634390fa66a4bfb998da691e7cd81cf45e38c6db. Software Heritage. https://archive.softwareheritage.org/swh:1:dir:4d4436affc6aa5ddf265a9a9acb638f66eb0cd5d;origin=https://github.com/timktsang/covid19_transmission_heterogeneity;visit=swh:1:snp:6e5eab4d82393ee3420ab8827a701a0129569d57;anchor=swh:1:rev:634390fa66a4bfb998da691e7cd81cf45e38c6db
- Tsang TK, Fang LQ, Zhang A, Jiang FC, Ruan SM, Liu LZ, Cowling BJ, Liu W, Yang Y.** 2022. Variability in transmission risk of SARS-cov-2 in close contact settings: A contact tracing study in shandong province, china. *Epidemics* **39**:100553. DOI: <https://doi.org/10.1016/j.epidem.2022.100553>
- Veroniki AA, Jackson D, Bender R, Kuss O, Langan D, Higgins JPT, Knapp G, Salanti G.** 2019. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Research Synthesis Methods* **10**:23–43. DOI: <https://doi.org/10.1002/jrsm.1319>, PMID: 30129707
- Wilkinson K, Chen X, Shaw S.** 2021. Secondary attack rate of COVID-19 in household contacts in the winnipeg health region, canada. *Canadian Journal of Public Health = Revue Canadienne de Sante Publique* **112**:12–16. DOI: <https://doi.org/10.17269/s41997-020-00451-x>, PMID: 33205377
- Wong G, Liu W, Liu Y, Zhou B, Bi Y, Gao GF.** 2015. MERS, SARS, and ebola: the role of super-spreaders in infectious disease. *Cell Host & Microbe* **18**:398–401. DOI: <https://doi.org/10.1016/j.chom.2015.09.013>, PMID: 26468744
- Wong F, Collins JJ.** 2020. Evidence that coronavirus superspreading is fat-tailed. *PNAS* **117**:29416–29418. DOI: <https://doi.org/10.1073/pnas.2018490117>, PMID: 33139561

- Xu CLH**, Raval M, Schnall JA, Kwong JC, Holmes NE. 2020. Duration of respiratory and gastrointestinal viral shedding in children with SARS-cov-2: A systematic review and synthesis of data. *The Pediatric Infectious Disease Journal* **39**:e249–e256. DOI: <https://doi.org/10.1097/INF.0000000000002814>, PMID: 32618932
- Yan D**, Zhang X, Chen C, Jiang D, Liu X, Zhou Y, Huang C, Zhou Y, Guan Z, Ding C, Chen L, Lan L, Fu X, Wu J, Li L, Yang S. 2021. Characteristics of viral shedding time in SARS-cov-2 infections: A systematic review and meta-analysis. *Frontiers in Public Health* **9**:652842. DOI: <https://doi.org/10.3389/fpubh.2021.652842>, PMID: 33816427
- Yang B**, Wu P, Lau EHY, Wong JY, Ho F, Gao H, Xiao J, Adam DC, Ng TWY, Quan J, Tsang TK, Liao Q, Cowling BJ, Leung GM. 2021. Changing disparities in coronavirus disease 2019 (COVID-19) burden in the ethnically homogeneous population of hong kong through pandemic waves: an observational study. *Clinical Infectious Diseases* **73**:2298–2305. DOI: <https://doi.org/10.1093/cid/ciab002>, PMID: 33406238
- Zhang Y**, Cen M, Hu M, Du L, Hu W, Kim JJ, Dai N. 2021. Prevalence and persistent shedding of fecal SARS-cov-2 RNA in patients with COVID-19 infection: A systematic review and meta-analysis. *Clinical and Translational Gastroenterology* **12**:e00343. DOI: <https://doi.org/10.14309/ctg.0000000000000343>, PMID: 33835096
- Zhao S**, Shen M, Musa SS, Guo Z, Ran J, Peng Z, Zhao Y, Chong MKC, He D, Wang MH. 2021. Inferencing superspreading potential using zero-truncated negative binomial model: exemplification with COVID-19. *BMC Medical Research Methodology* **21**:30. DOI: <https://doi.org/10.1186/s12874-021-01225-w>, PMID: 33568100

Appendix 1

Systematic review

We conducted a systematic review following the Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) statement (**Moher et al., 2009**). A standardized search was done in PubMed, Embase, and Web of Science, using the search term "(“COVID-19” OR “SARS-CoV-2”) AND (“household” OR “family”) AND (“transmission”)". The search was done on 22 June, 2022, with no language restrictions. Additional relevant articles from the reference sections were also reviewed.

Two authors (XH and CW) independently screened the titles and extracted data from the included studies, with disagreement resolved by consensus together with a third author (TKT). Studies identified from different databases were de-duplicated.

Household transmission model

Our aim was to explore the role of variation of individual infectiousness on transmission, with accounting for the variation and number of contacts, and the transmission probability. Therefore, we extended the household transmission model to account for variation of individual infectiousness of index cases (**Cauchemez et al., 2009**). We added a random effect on infectiousness of cases (**Tsang et al., 2014**). In the model, the hazard of infection of individual j at time t from an infected household member i , with infection time t_i in household k , was

$$\lambda_{i \rightarrow j}(t) = \frac{\lambda_h}{X_k^\beta} * \exp(\delta_i) * f(t - t_i)$$

where λ_h was the baseline hazard, δ_i was the relative individual infectiousness of cases, which followed a normal distribution with mean 0 and SD σ_{var} .

X_k was the number of household contacts. β was the parameter describing the relationship between number of household contacts and transmission rate. It ranged from 0 to 1, with 0 indicating that the transmission rate is independent of number of household contacts while 1 indicates that the transmission rate is inversely proportional to number of household contacts (i.e. dilution effect of the contact time per contact which is lower when the number of household contacts is higher). $f(\cdot)$ was the infectiousness profile since infection generated from the assumed incubation period (mean equal to 5 days) and infectious period (mean equal to 13) (**He et al., 2020; Jing et al., 2020; Li et al., 2020b**).

The daily hazard of infection from outside the household is assumed to be $\lambda_c(t) = \lambda_c$. Hence, the total hazard infection of individual j on day t was

$$\lambda_j(t) = \lambda_c + \sum_{i: \text{infected}} \lambda_{i \rightarrow j}(t)$$

Based on the transmission model, the probability that an individual i was infected with infection time t_{i0} was

$$P(y_i = 1, t_i = t_{i0}) = [1 - \exp(-\lambda_i(t_{i0}))] * \left[\exp\left(-\sum_{d=z_{i1}}^{t_{i0}-1} \lambda_i(d)\right) \right]$$

where z_{i1} was the start of the follow-up day of individual i . Denote the total follow-up time for an individual i as z_{i2} , then the probability that an individual i did not get infected within the follow-up period was

$$P(y_i = 0, t_i = t_{i0} = z_{i2} + 1) = \exp\left(-\sum_{d=z_{i1}}^{t_{i0}-1} \lambda_i(d)\right)$$

Hence the log-likelihood function L was

$$L = \sum_{i: y_i=1} \log(1 - \exp(-\lambda_i(t_{i0}))) - \sum_i \sum_{d=z_{i1}}^{t_{i0}-1} \lambda_i(d)$$

Index cases did not contribute to the likelihood and hence the summation was only on household contacts.

Inference, model comparison, and validation

The infection time for cases were unavailable in the data extracted from publication. Also, the individual infectiousness parameter δ_i was an augmented variable. Therefore, we conducted our inference under a Bayesian framework with using data augmentation MCMC algorithm to joint estimate the model parameters, the missing infection time, and the relative individual infectiousness with using metropolis-hasting algorithm. For the parameter that only takes positive value, we used a vague Uniform(0,10) prior. For the parameter for relationship between number of household contacts and transmission, we used Uniform(0,1). The algorithm runs for 50,000 iterations after a burn-in of 50,000 iterations. Convergence was visually assessed. One run takes about 3 hr on our typical desktop computer. We fitted the model to all studies, but only 14 of them could converge and provide robust estimates of individual infectiousness.

We assessed the model adequacy by comparing the observed and expected number of infections in households by household size. We simulated 10,000 datasets with a structure identical to that of the observed data (in terms of number of household contacts), with parameters randomly drawn from the posterior distribution of model parameter.

We compared the model with or without the random effect for variation of individual infectiousness by using DIC (*Spiegelhalter et al., 2002*). Smaller DIC indicated a better model fit. DIC difference >5 was considered as substantial improvement (*Spiegelhalter et al., 2000*). Since the likelihood of observed data could not be directly evaluated for a given model (*Celeux et al., 2006*), we used an importance sampling approach to estimate the likelihood for the observed data and evaluate the DIC (*Cauchemez et al., 2014; Liu, 2001*). For each household, we simulated 2000 datasets, with parameters randomly drawn from the posterior distribution. Then we compared the observed data and simulated data. The contribution to the likelihood of a household was equal to the proportion of simulated data with infection status that exactly matched the observed data, for all household members. To avoid the problem of 0-valued likelihood, we used the approach developed by *Cauchemez et al., 2014*, and assumed the sensitivity and specificity for diagnosing a case were both 99.99%.

Comparison of modeling approach

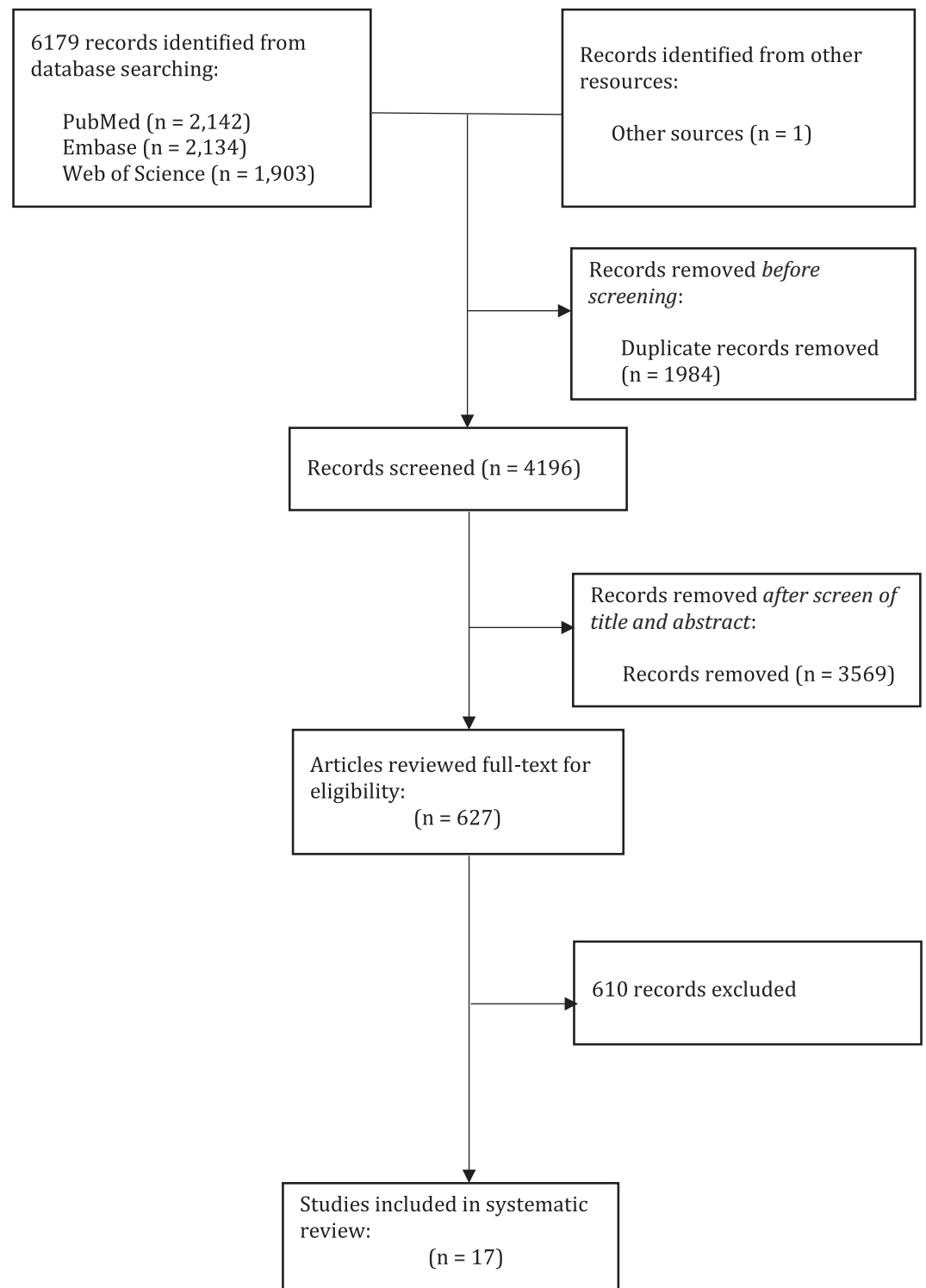
Given that the infection time (or information that could inform infection time, such as symptom onset time) were not available for all studies, it may be possible to consider them as final size data, which may be analyzed by chain binomial model {Longini, 1988 #693; Longini, 1982 #335; Klick, 2011 #338; O'Neill, 2000 #368}. Including covariates affecting susceptibility is still possible, but showed to be difficult since the number of parameters increases exponentially by groups and type of infection {Klick, 2011 #338}. However, the model equation is about the escape probability of infection, which did not include the characteristic of infectee and therefore it is not possible to add covariates affecting infectiousness.

Sensitivity analysis

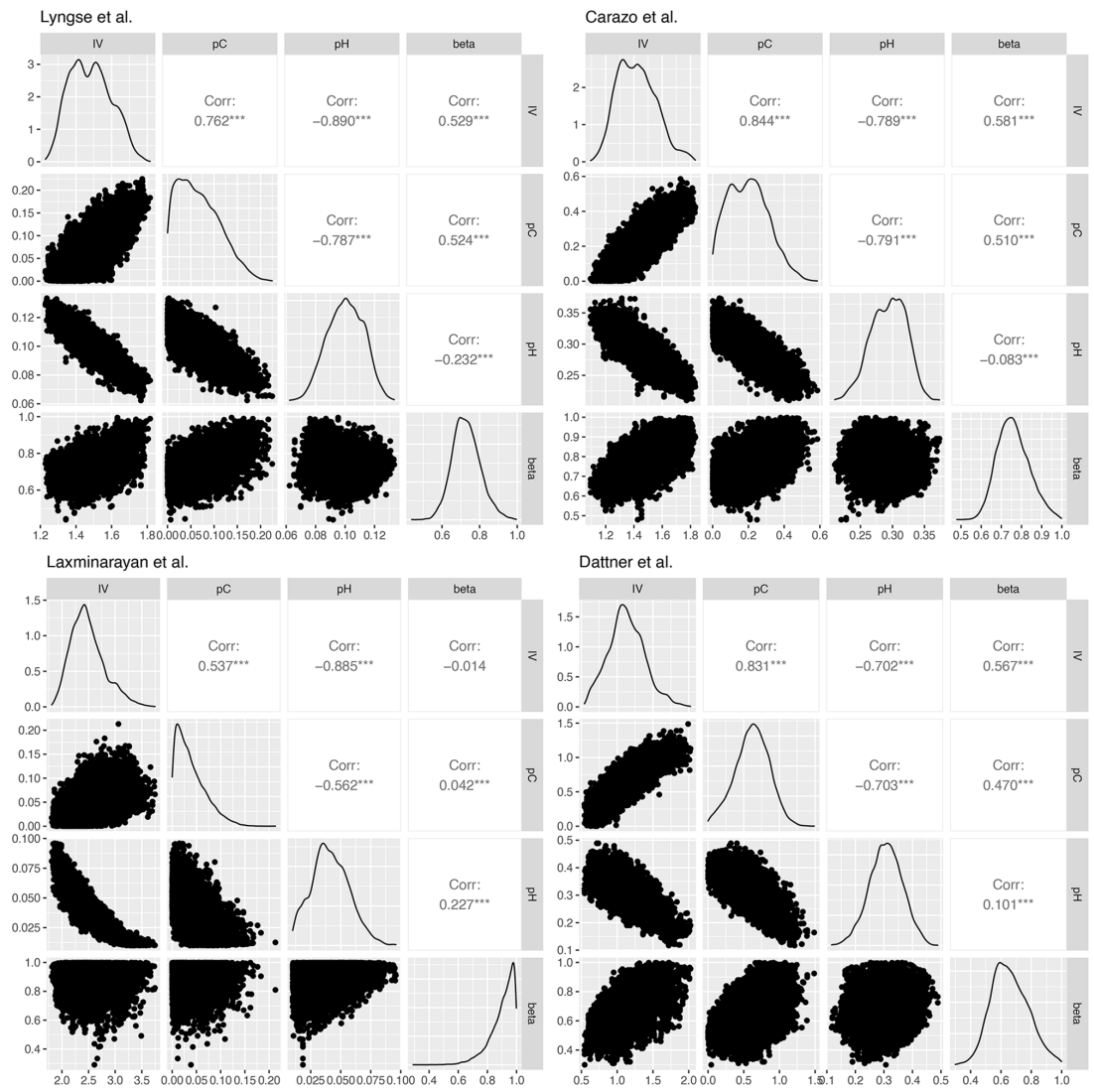
Our model assumed that the individual infectiousness parameter δ_i followed a normal distribution, and it was multiplied to the hazard in the form of $\exp(\delta_i)$, therefore, the individual infectiousness essentially followed a lognormal distribution, which had a long tail. Therefore, we conducted a sensitivity analysis as follows:

$$\lambda_{i \rightarrow j}(t) = \frac{\lambda_j}{X_k^\beta} * \delta_i * f(t - t_i)$$

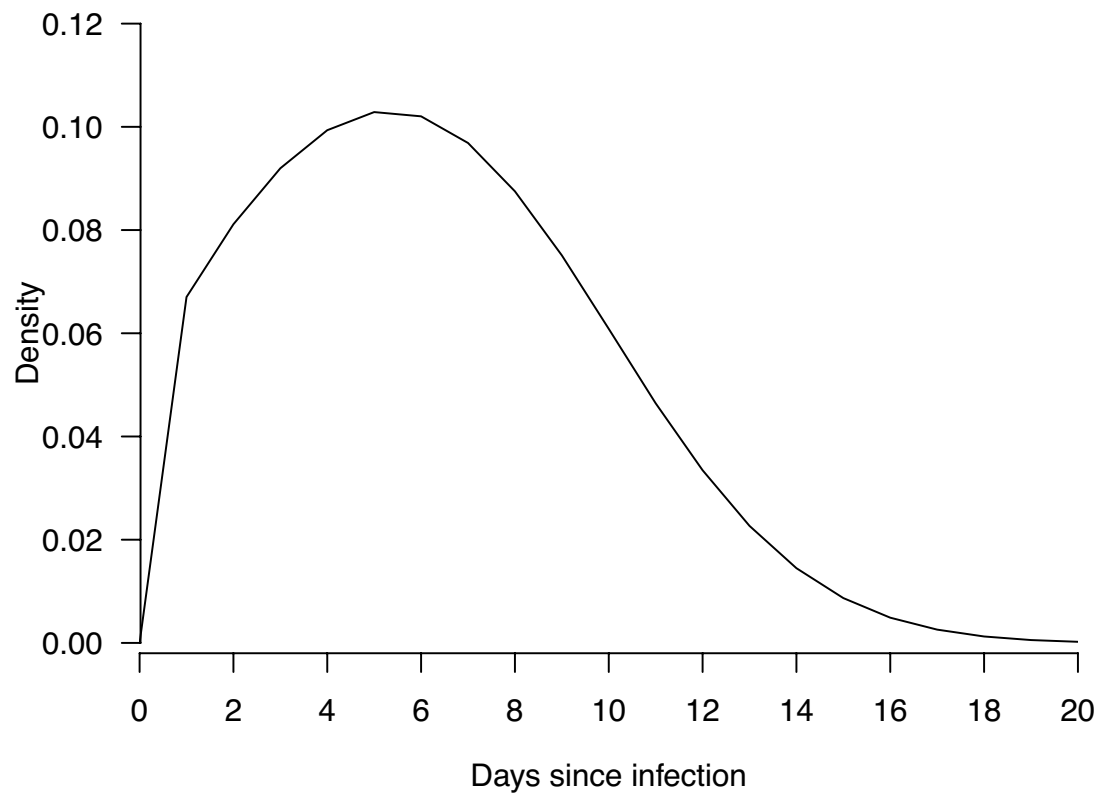
Here, δ_i followed a gamma distribution with mean 1, and SD σ_{var} , which had a shorter tail compared with lognormal distribution. Model comparison suggested that assuming gamma distribution for individual infectiousness parameter performed substantially worse, compared with the lognormal distribution in the main analysis (**Appendix 1—table 9**).



Appendix 1—figure 1. Process of systematic review.



Appendix 1—figure 2. Correlation plots for the posterior distribution of model parameters in Lyngse et al., Carazo et al., Laxminarayan et al., and Dattner et al.



Appendix 1—figure 3. The density plot of the distribution of infectiousness profile since infections, used in the modeling analysis.

Appendix 1—table 1. Summary of characteristic of identified studies.

Author (year)	Location	Study period	Case ascertainment method	Test coverage of identified contacts	SARS-COV-2 variant	Public health and social measures
<i>Lyngse et al., 2022</i>	Denmark	February 2020 to August 2020	Index: RT-PCR Secondary: RT-PCR	All contacts were tested	Ancestral strains	Lockdown
<i>Carazo et al., 2022</i>	Canada	March 2020 to June 2020	Index: RT-PCR Secondary: Symptom-based	NA	Ancestral strains	Handwashing, mask-wearing, and physical distancing
<i>Laxminarayan et al., 2020</i>	India	March 2020 to July 2020	Index: RT-PCR Secondary: RT-PCR	All contacts were tested	Ancestral strains	Lockdown, social distancing, contact tracing
<i>Dattner et al., 2021</i>	Israel	March 2020 to June 2020	Index: RT-PCR and Serology Secondary: RT-PCR and Serology	All contacts were tested	Ancestral strains	Lockdown
<i>Layan et al., 2022</i>	Israel	December 2020 to April 2021	Index: RT-PCR Secondary: RT-PCR	All contacts were tested	Ancestral strains, alpha	Vaccination
<i>Hart et al., 2022</i>	UK	March 2020 to November 2020	Index: RT-PCR Secondary: RT-PCR and antibody test	All contacts were tested	Ancestral strains	Isolation
<i>Hubiche et al., 2021</i>	Canada	April 2020 to June 2020	Index: Symptom-based and RT-PCR and serology Secondary: Symptom-based and RT-PCR and serology	NA	Ancestral strains	Closure of school
<i>Wilkinson et al., 2021</i>	Manitoba, Canada	Mid-January 2020 to late March 2020	Index: NAAT assay Secondary: NAAT assay	Only symptomatic contacts were tested	Ancestral strains	Isolation and contact tracing

Appendix 1—table 1 Continued on next page

Appendix 1—table 1 Continued

Author (year)	Location	Study period	Case ascertainment method	Test coverage of identified contacts	SARS-COV-2 variant	Public health and social measures
Tsang et al., 2022	Shandong Province, China	January 2020 to May 2020	Index: RT-PCR Secondary: RT-PCR	All contacts were tested	Ancestral strains	Isolation, mask-wearing, social distancing
Reukers et al., 2021	Netherlands	March 2020 to May 2020	Index: RT-PCR Secondary: RT-PCR	NA	Ancestral strains	Social distancing
Han et al., 2022	Netherlands	March 2020 to April 2020	Index: RT-PCR and serology Secondary: RT-PCR and serology	All contacts were tested	Ancestral strains	Social distancing, self-quarantine and self-isolation orders, closure of schools, bars and restaurants, and urging people to work from home
Méndez-Echevarría et al., 2021	Madrid, Spain	March 2020 to May 2020	Index: RT-PCR and serology Secondary: RT-PCR and serology	All contacts were tested	Ancestral strains	Lockdown
Dutta et al., 2020	Rajasthan, India	May 2020 to July 2020	Index: RT-PCR Secondary: RT-PCR	All contacts were tested	Ancestral strains	Lockdown and stay-at-home orders, physical distancing
Koureas et al., 2021	Greece	April 2020 to June 2020	Index: RT-PCR Secondary: RT-PCR	All contacts were tested	Ancestral strains	Quarantine, screening, movement restrictions and gathering prohibition, isolation
Bernardes-Souza et al., 2021; Méndez-Echevarría et al., 2021	Brazil	May 2020 to June 2020	Index: Serology and RT-PCR Secondary: Serology	All contacts were tested	Ancestral strains	Lockdown, gathering restrictions and face mask mandates
Posfay-Barbe et al., 2020	Switzerland	March 2020 to April 2020	Index: RT-PCR Secondary: RT-PCR	NA	Ancestral strains	Closure of schools, daycares, restaurants, bars, and shops, social distancing
Hsu et al., 2021	Taiwan, China	January 2020 to February 2021	Index: RT-PCR Secondary: RT-PCR	Only symptomatic contacts were tested	Ancestral strains, alpha	Mask-wearing, hand washing, and social distancing

Appendix 1—table 2. Summary of model estimates.

Article	Estimates of infectiousness variation	Estimates of probability of infection from community (10^{-2})	Estimates of probability of infection from household	Relationship between transmission and number of contacts (β)
Lyngse et al.	1.48 (1.29, 1.7)	0.06 (0, 0.16)	0.1 (0.08, 0.12)	0.72 (0.59, 0.89)
Carazo et al.	1.41 (1.19, 1.72)	0.2 (0.01, 0.43)	0.3 (0.24, 0.34)	0.75 (0.62, 0.93)
Laxminarayan et al.	2.44 (1.98, 3.23)	0.03 (0, 0.11)	0.04 (0.01, 0.07)	0.92 (0.69, 1)
Dattner et al.	1.12 (0.65, 1.69)	0.63 (0.12, 1.06)	0.31 (0.19, 0.42)	0.65 (0.45, 0.91)
Layan et al.	1.12 (0.74, 1.76)	0.3 (0.02, 0.96)	0.26 (0.15, 0.39)	0.2 (0.01, 0.6)
Hart et al.	0.35 (0.11, 0.96)	0.74 (0.05, 2.12)	0.51 (0.32, 0.65)	0.72 (0.32, 0.98)
Hubiche et al.	1.03 (0.45, 2.68)	1.24 (0.14, 2.48)	0.32 (0.06, 0.57)	0.67 (0.13, 0.98)
Wilkinson et al.	1.05 (0.12, 3.51)	0.32 (0.02, 0.91)	0.07 (0, 0.18)	0.43 (0.02, 0.97)
Tsang et al.	2.83 (1.48, 4.73)	0.45 (0.05, 1.03)	0.06 (0, 0.26)	0.66 (0.06, 0.99)
Reukers et al.	1.79 (0.72, 4.5)	1.53 (0.18, 2.93)	0.22 (0.01, 0.6)	0.67 (0.07, 0.98)
Han et al.	1.71 (0.51, 4.13)	1.67 (0.19, 2.99)	0.21 (0.02, 0.61)	0.69 (0.07, 0.99)
Méndez-Echevarría et al.	0.51 (0.13, 1.7)	0.91 (0.05, 2.53)	0.28 (0.06, 0.52)	0.57 (0.04, 0.98)
Dutta et al.	1.53 (0.5, 4.2)	1.1 (0.12, 2.26)	0.24 (0.01, 0.63)	0.78 (0.15, 0.99)

Appendix 1—table 2 Continued on next page

Appendix 1—table 2 Continued

Article	Estimates of infectiousness variation	Estimates of probability of infection from community (10^{-2})	Estimates of probability of infection from household	Relationship between transmission and number of contacts (β)
Koureas et al.	1.88 (0.6, 3.66)	0.84 (0.1, 1.76)	0.13 (0.01, 0.42)	0.44 (0.02, 0.96)

Appendix 1—table 3. Model adequacy check for Lyngse et al.

Each element of the table has the format observed frequency – expected (posterior mean) frequency (95% credible interval).

Number of household contacts	Number of infected household contacts					
	0	1	2	3	4	5
1	2366–2377 (2319, 2433)	569–558 (502, 616)	NA	NA	NA	NA
2	1117–1105 (1070, 1138)	227–227 (198, 259)	77–88 (69, 109)	NA	NA	NA
3	1135–1118 (1077, 1156)	214–235 (203, 267)	89–86 (67, 106)	41–41 (27, 56)	NA	NA
4	521–528 (501, 555)	119–115 (94, 137)	40–42 (30, 56)	25–19 (11, 30)	11–10 (4, 18)	NA
5	161–167 (152, 181)	42–38 (27, 50)	14–14 (7, 22)	7–7 (2, 12)	7–3 (0, 8)	0–2 (0, 5)

Appendix 1—table 4. Model adequacy check for Carazo et al.

Each element of the table has the format observed frequency – expected (posterior mean) frequency (95% credible interval).

Number of household contacts	Number of infected household contacts					
	0	1	2	3	4	5
1	803–814 (765, 862)	532–521 (473, 570)	NA	NA	NA	NA
2	476–454 (423, 485)	179–186 (160, 212)	154–169 (144, 195)	NA	NA	NA
3	518–520 (483, 557)	201–202 (175, 232)	132–130 (107, 155)	133–131 (106, 158)	NA	NA
4	217–221 (197, 245)	85–87 (70, 106)	52–52 (39, 67)	43–39 (27, 52)	45–41 (28, 56)	NA
5	76–74 (61, 88)	30–30 (20, 41)	20–18 (10, 26)	13–12 (6, 20)	7–10 (5, 17)	10–11 (5, 18)

Appendix 1—table 5. Model adequacy check for Laxminarayan et al.

Each element of the table has the format observed frequency – expected (posterior mean) frequency (95% credible interval).

Number of household contacts	Number of infected household contacts									
	0	1	2	3	4	5	6	7	8	9
1	124–134 (122, 144)	37–27 (17, 39)	NA	NA	NA	NA	NA	NA	NA	NA
2	135–137 (125, 147)	12–18 (10, 27)	20–12 (6, 20)	NA	NA	NA	NA	NA	NA	NA
3	188–176 (163, 188)	13–22 (13, 33)	11–9 (4, 16)	5–9 (4, 17)	NA	NA	NA	NA	NA	NA
4	127–118 (108, 128)	11–15 (8, 23)	5–5 (1, 11)	3–3 (0, 7)	1–4 (1, 10)	NA	NA	NA	NA	NA
5	75–76 (67, 84)	12–10 (4, 16)	3–3 (0, 8)	1–2 (0, 5)	1–1 (0, 4)	3–2 (0, 6)	NA	NA	NA	NA
6	41–40 (34, 46)	7–5 (1, 10)	2–2 (0, 5)	0–1 (0, 3)	0–0 (0, 3)	0–0 (0, 2)	1–1 (0, 3)	NA	NA	NA
7	32–31 (25, 36)	3–4 (1, 9)	2–1 (0, 4)	1–1 (0, 3)	1–0 (0, 2)	0–0 (0, 2)	0–0 (0, 2)	0–0 (0, 3)	NA	NA
8	10–13 (10, 16)	2–2 (0, 5)	3–0 (0, 2)	1–0 (0, 2)	0–0 (0, 1)	1–0 (0, 1)	0–0 (0, 1)	0–0 (0, 1)	0–0 (0, 1)	NA
9	14–16 (12, 20)	2–2 (0, 6)	1–1 (0, 3)	2–0 (0, 2)	1–0 (0, 2)	0–0 (0, 1)	0–0 (0, 1)	0–0 (0, 1)	0–0 (0, 1)	1–0 (0, 2)

Appendix 1—table 6. Model adequacy check for Dattner et al.

Each element of the table has the format observed frequency – expected (posterior mean) frequency (95% credible interval).

Number of household contacts	Number of infected household contacts									
	0	1	2	3	4	5	6	7	8	9
1	85–93 (77, 108)	73–65 (50, 81)	NA	NA	NA	NA	NA	NA	NA	NA
2	46–42 (32, 52)	21–25 (17, 34)	19–19 (11, 28)	NA	NA	NA	NA	NA	NA	NA
3	38–29 (21, 38)	13–18 (11, 25)	9–12 (6, 19)	8–9 (4, 15)	NA	NA	NA	NA	NA	NA
4	24–24 (16, 32)	19–15 (9, 23)	11–10 (5, 17)	8–7 (3, 13)	1–5 (1, 11)	NA	NA	NA	NA	NA
5	25–21 (14, 30)	15–15 (9, 22)	12–10 (5, 17)	7–7 (3, 13)	3–5 (2, 10)	2–4 (1, 9)	NA	NA	NA	NA
6	13–16 (9, 23)	19–12 (6, 18)	3–8 (3, 14)	5–6 (2, 11)	3–5 (1, 9)	5–3 (0, 7)	5–2 (0, 6)	NA	NA	NA
7	5–9 (4, 15)	6–7 (3, 13)	5–5 (2, 10)	7–4 (1, 8)	4–3 (0, 7)	4–2 (0, 6)	4–2 (0, 5)	0–1 (0, 4)	NA	NA
8	10–8 (3, 14)	7–7 (3, 12)	3–5 (1, 10)	4–4 (1, 8)	0–3 (0, 7)	3–2 (0, 6)	3–2 (0, 5)	4–1 (0, 4)	0–1 (0, 4)	NA
9	6–6 (2, 12)	3–6 (2, 10)	4–4 (1, 9)	6–3 (0, 7)	4–3 (0, 6)	1–2 (0, 5)	2–2 (0, 5)	2–1 (0, 4)	2–1 (0, 3)	0–0 (0, 3)

Appendix 1—table 7. Comparison of model estimates from 100k Markov chain Monte Carlo (MCMC) iterations and 500k MCMC iterations.

Article	Number of MCMC iterations	Estimates of infectiousness variation	Estimates of probability of infection from community (10^{-2})	Estimates of probability of infection from household	Relationship between transmission and number of contacts (β)
Lyngse et al.	100,000	1.48 (1.29, 1.7)	0.06 (0, 0.16)	0.11 (0.08, 0.13)	0.72 (0.59, 0.89)
	500,000	1.49 (1.29, 1.84)	0.07 (0, 0.2)	0.1 (0.07, 0.13)	0.74 (0.6, 0.92)
Carazo et al.	100,000	1.41 (1.19, 1.72)	0.2 (0.01, 0.43)	0.35 (0.27, 0.42)	0.75 (0.62, 0.93)
	500,000	1.44 (1.19, 1.78)	0.21 (0.02, 0.46)	0.34 (0.26, 0.41)	0.77 (0.62, 0.95)
Laxminarayan et al.	100,000	2.44 (1.98, 3.23)	0.03 (0, 0.11)	0.04 (0.01, 0.08)	0.92 (0.69, 1)
	500,000	2.47 (1.99, 3.23)	0.03 (0, 0.11)	0.04 (0.01, 0.07)	0.92 (0.67, 1)
Dattner et al.	100,000	1.12 (0.65, 1.69)	0.63 (0.12, 1.06)	0.37 (0.21, 0.54)	0.65 (0.45, 0.91)
	500,000	1.06 (0.65, 1.55)	0.57 (0.1, 1)	0.38 (0.24, 0.54)	0.63 (0.43, 0.88)

Appendix 1—table 8. Simulation study for validating the estimates and the corresponding power.

Parameter	Simulation value	Mean estimate	Proportion covered (over 50)
Lyngse et al.			
σ_{var} : Infectiousness variation	1.48	1.57	0.94
λ_c : hazard of infection from outside the household (10^{-2})	0.06	0.12	0.88
λ_h : hazard of infection from an infected household member	0.11	0.09	0.92
β : relationship between number of household contacts and transmission rate	0.72	0.75	0.96
Carazo et al.			
σ_{var} : Infectiousness variation	1.41	1.49	0.94
λ_c : hazard of infection from outside the household (10^{-2})	0.2	0.24	0.96
λ_h : hazard of infection from an infected household member	0.35	0.33	0.92
β : relationship between number of household contacts and transmission rate	0.75	0.77	0.96

Appendix 1—table 8 Continued on next page

Appendix 1—table 8 Continued

Parameter	Simulation value	Mean estimate	Proportion covered (over 50)
Laxminarayan et al.			
σ_{var} : Infectiousness variation	2.44	2.6	0.94
λ_c : hazard of infection from outside the household (10^{-2})	0.03	0.05	0.96
λ_h : hazard of infection from an infected household member	0.04	0.03	0.9
β : relationship between number of household contacts and transmission rate	0.92	0.82	1
Dattner et al.			
σ_{var} : Infectiousness variation	1.12	1.18	0.92
λ_c : hazard of infection from outside the household (10^{-2})	0.63	0.69	0.92
λ_h : hazard of infection from an infected household member	0.37	0.36	0.96
β : relationship between number of household contacts and transmission rate	0.65	0.68	0.98

Appendix 1—table 9. A sensitivity analysis of using normal distribution (main analysis) instead of lognormal distribution (sensitivity analysis) for individual infectiousness.

Article	Model	Estimates of infectiousness variation	Estimates of probability of infection from community (10^{-2})	Estimates of probability of infection from household	Relationship between transmission and number of contacts (β)	Difference in DIC
Lyngse et al.	Main analysis	1.48 (1.29, 1.7)	0.06 (0, 0.16)	0.11 (0.08, 0.13)	0.72 (0.59, 0.89)	
	Sensitivity analysis	0.93 (0.84, 1.03)	0.03 (0, 0.11)	0.06 (0.05, 0.07)	0.7 (0.58, 0.83)	2821
Carazo et al.	Main analysis	1.41 (1.19, 1.72)	0.2 (0.01, 0.43)	0.35 (0.27, 0.42)	0.75 (0.62, 0.93)	
	Sensitivity analysis	1.05 (0.92, 1.21)	0.06 (0, 0.24)	0.18 (0.16, 0.2)	0.71 (0.59, 0.84)	2569
Laxminarayan et al.	Main analysis	2.44 (1.98, 3.23)	0.03 (0, 0.11)	0.04 (0.01, 0.08)	0.92 (0.69, 1)	
	Sensitivity analysis	1.31 (1.14, 1.48)	0.02 (0, 0.08)	0.05 (0.03, 0.06)	0.91 (0.66, 1)	818
Dattner et al.	Main analysis	1.12 (0.65, 1.69)	0.63 (0.12, 1.06)	0.37 (0.21, 0.54)	0.65 (0.45, 0.91)	
	Sensitivity analysis	0.66 (0.47, 0.88)	0.3 (0.02, 0.71)	0.18 (0.13, 0.22)	0.56 (0.4, 0.75)	908

Appendix 1—table 10. Summary of characteristic of identified studies.

SD: standard deviation.

Article	Number of households	Number of contacts	Number of secondary cases	Mean number of contacts	SD of number of contact	Secondary attack rate (SAR)	SD of number of secondary cases (σ_{sec})
Lyngse et al.	6782	14,233	1902	2.1 (2.07, 2.13)	1.17 (1.15, 1.19)	0.13 (0.13, 0.14)	0.6 (0.59, 0.61)
Carazo et al.	3727	8460	2574	2.27 (2.23, 2.31)	1.19 (1.16, 1.21)	0.3 (0.29, 0.31)	0.97 (0.95, 0.99)
Laxminarayan et al.	915	3113	283	3.4 (3.28, 3.53)	1.94 (1.86, 2.04)	0.09 (0.08, 0.1)	0.81 (0.77, 0.85)
Dattner et al.	591	2211	720	3.74 (3.54, 3.94)	2.5 (2.36, 2.65)	0.33 (0.31, 0.35)	1.6 (1.51, 1.7)
Layan et al.	208	670	264	3.22 (3.01, 3.44)	1.58 (1.45, 1.75)	0.39 (0.36, 0.43)	1.59 (1.45, 1.76)
Hart et al.	172	433	194	2.52 (2.34, 2.69)	1.16 (1.05, 1.29)	0.45 (0.4, 0.5)	1.11 (1, 1.24)
Hubiche et al.	103	291	119	2.83 (2.6, 3.05)	1.18 (1.04, 1.37)	0.41 (0.35, 0.47)	1.14 (1, 1.32)
Wilkinson et al.	95	220	26	2.32 (2.02, 2.61)	1.47 (1.28, 1.71)	0.12 (0.08, 0.17)	0.66 (0.58, 0.77)

Appendix 1—table 10 Continued on next page

Appendix 1—table 10 Continued

Article	Number of households	Number of contacts	Number of secondary cases	Mean number of contacts	SD of number of contact	Secondary attack rate (SAR)	SD of number of secondary cases (σ_{sec})
Tsang et al.	47	189	38	4.02 (3.39, 4.66)	2.22 (1.85, 2.79)	0.2 (0.15, 0.27)	1.31 (1.09, 1.65)
Reukers et al.	55	187	78	3.4 (3.11, 3.69)	1.1 (0.93, 1.35)	0.42 (0.35, 0.49)	1.33 (1.12, 1.64)
Han et al.	55	185	78	3.36 (3.08, 3.65)	1.08 (0.91, 1.33)	0.42 (0.35, 0.5)	1.27 (1.07, 1.57)
Méndez-Echevarría et al.	63	174	57	2.76 (2.57, 2.95)	0.78 (0.66, 0.94)	0.33 (0.26, 0.4)	0.96 (0.82, 1.17)
Dutta et al.	32	170	55	5.31 (4.52, 6.1)	2.28 (1.83, 3.03)	0.32 (0.25, 0.4)	1.59 (1.28, 2.12)
Koureas et al.	32	153	50	4.78 (3.94, 5.62)	2.43 (1.95, 3.23)	0.33 (0.25, 0.41)	2.14 (1.72, 2.84)
Bernardes-Souza et al.	40	112	55	2.8 (2.33, 3.27)	1.51 (1.23, 1.93)	0.49 (0.4, 0.59)	1.48 (1.21, 1.9)
Posfay-Barbe et al.	39	111	46	2.85 (2.54, 3.16)	0.99 (0.81, 1.27)	0.41 (0.32, 0.51)	0.94 (0.77, 1.21)
Hsu et al.	38	96	49	2.53 (2.12, 2.93)	1.27 (1.03, 1.64)	0.51 (0.41, 0.61)	0.87 (0.71, 1.12)

Appendix 1—table 11. Association between infectiousness variation estimated from household transmission models, and other statistics from 17 household studies, based on meta-regression.

Statistic	Infectiousness variation (σ_{var}^*)	Standard deviation (SD) of the distribution of number of secondary cases (σ_{sec})	Secondary attack rate (SAR)
Pooled estimates	1.33 (0.95, 1.70)	1.19 (1.03, 1.35)	0.35 (0.28, 0.44)
I^2	78.4	99.2	99.1
Factors			
Estimate	β	β	exp (β)
Mean number of contacts	0.31 (−0.16, 0.78)	0.35 (0.22, 0.48)	1.09 (0.83, 1.43)
SD number of contacts	0.55 (−0.12, 1.21)	0.42 (0.15, 0.69)	0.80 (0.51, 1.25)
Implementation of lockdown	0.16 (−0.61, 0.93)	−0.06 (−0.40, 0.28)	0.69 (0.43, 1.10)
Index cases are confirmed by PCR only	0.48 (−0.32, 1.29)	0.02 (−0.33, 0.38)	1.09 (0.65, 1.85)
Secondary cases are confirmed by PCR only	0.78 (0.13, 1.43)	0.03 (−0.30, 0.35)	0.88 (0.55, 1.41)
Only ancestral strains are circulating in study period	0.76 (−0.03, 1.55)	0.07 (−0.31, 0.45)	0.75 (0.43, 1.30)
All household contacts were tested	0.03 (−0.87, 0.93)	0.28 (−0.04, 0.59)	0.96 (0.58, 1.58)

*Estimates based on results from 14 studies

Appendix 1—table 12. Pooled estimates for duration of viral shedding for COVID-19 from 11 systematic reviews.

Study	Sampling site	Laboratory method	Pooled estimates (mean/median)	95% CI	Range	I^2
Cevik	Upper respiratory tract	PCR	17.0	15.5–18.6	6.0–53.9	
Cevik	Lower respiratory tract	PCR	14.6	9.3–20.0	6.2–22.7	97%
Cevik	Stool	Viral culture	17.2	14.4–20.1	9.8–27.9	96.6%
Cevik	Serum/blood	Viral culture	16.6	3.6–29.7	10.0–23.3	99%
Fontana	Respiratory sources	PCR	18.4*	15.5–21.3	5.5–53.5	98.87%
Fontana	Rectal/stool	PCR	22.1*	14.4–29.8	11–33	95.86%

Appendix 1—table 12 Continued on next page

Appendix 1—table 12 Continued

Study	Sampling site	Laboratory method	Pooled estimates (mean/median)	95% CI	Range	I ²
Okita	Upper respiratory tract (nasal swab+throat swab)	PCR	18.29	17.00–19.89	8.33–39.97	99%
Okita	Sputum	PCR	23.79	20.43–27.16	15.50–32.00	93%
Okita	Blood	PCR	14.60	12.16–17.05	11.00–17.58	88%
Okita	Stool	PCR	22.38	18.40–26.35	10.67–51.40	97%
Qutub	Respiratory tract	Viral culture	28.75/11**	8.5–14.5		
Rahmani	Respiratory tract	PCR	27.90	23.27–32.52	7.40–132.00	99.1%
Xu	Respiratory tract (symptomatic cases)	PCR	11.1±5.8***		0–24	
Xu	Gastrointestinal tract (symptomatic cases)	PCR	23.6±8.8***		10–33	
Yan	Unrestricted	PCR	16.8	14.8–19.4		99.56%
Yan	Stool	PCR	30.3	23.1–39.2		92.09%
Yan	Respiratory tract	PCR	17.5	14.9–20.6		99.67%
Yan	Upper respiratory tract	PCR	17.5	14.6–21.0		99.53%
Diaz	Stool	PCR	22*	19–25		
Chen	(Asymptomatic infections)		14.14*	11.25–17.04	11.00–17.25	
Li	Upper respiratory tract (nasopharyngeal/throat swabs)	PCR	11.43	10.1–12.77		75.3%
Zhang	Stool	PCR	21.8	16.4–27.1		
Zhang	Respiratory tract	PCR	14.7	9.9–19.5		

*Median estimate, **median/grouped median, ***this study analyzed by cases, and reported mean ± SD.

Appendix 1—table 13. Pooled estimates for duration of viral shedding for COVID-19 in subgroups from seven systematic reviews.

Study	Sampling site (subgroups)	Laboratory method	Pooled estimates (mean/median)	95% CI	Range	I ²
Fontana	Respiratory sources (among severely ill patients)	PCR	19.8*	16.2–23.5	11–31	96.42%
Fontana	Respiratory sources (in mild-to-moderate illness)	PCR	17.2*	14.0–20.5	8–25	95.64%
Okita	Upper respiratory tract (nasal swab)	PCR	19.34	16.60–22.07		99%
Okita	Upper respiratory tract (throat swab)	PCR	17.85	16.43–19.26		98%
Okita	Upper respiratory tract (age < 60)	PCR	16.95	13.56–20.35	8.62–35.67	98%
Okita	Upper respiratory tract (age ≥ 60)	PCR	21.24	14.06–28.41	8.25–40.33	99%
Okita	Upper respiratory tract (with comorbidities)	PCR	20.26	17.60–22.92	9.67–34.00	93%
Okita	Upper respiratory tract (without comorbidities)	PCR	14.66	12.63–16.69	10.82–27.25	85%
Okita	Upper respiratory tract (severe patients)	PCR	20.79	18.03–23.55	14.00–38.33	98%
Okita	Upper respiratory tract (nonsevere patients)	PCR	16.36	14.07–18.66	8.00–37.33	99%

Appendix 1—table 13 Continued on next page

Appendix 1—table 13 Continued

Study	Sampling site (subgroups)	Laboratory method	Pooled estimates (mean/median)	95% CI	Range	I ²
Okita	Upper respiratory tract (severe patients) (for studies with mean age ≥ 40 + comorbidity > 30%)	PCR	21.53	17.57–25.50	14.00–29.50	91%
Okita	Upper respiratory tract (severe patients) (for studies with mean age ≥ 40 + comorbidity >30%)	PCR	20.08	15.87–24.29	13.12–33.67	100%
Okita	Upper respiratory tract (treated with glucocorticoid)	PCR	19.72	17.92–21.52	13.87–33.67	92%
Okita	Upper respiratory tract (treated without glucocorticoid)	PCR	15.64	14.18–17.10	8.33–31.60	96%
Okita	Upper respiratory tract (treated with glucocorticoid) (for studies with mean age: 30–60+comorbidity > 50%)	PCR	21.98	16.48–27.48	14.25–33.67	94%
Okita	Upper respiratory tract (treated without glucocorticoid) (for studies with mean age: 30–60+comorbidity > 50%)	PCR	16.14	12.60–19.68	13.22–24.44	92%
Okita	Upper respiratory tract (Asian)	PCR	18.10	16.95–19.25	8.33–34.75	98%
Okita	Upper respiratory tract (European)	PCR	19.27	11.59–26.95	8.50–39.97	100%
Okita	Upper respiratory tract (Asian) (for studies with mean age ≥ 40 + comorbidity >40%)	PCR	20.66	18.18–23.14	12.00–32.00	96%
Okita	Upper respiratory tract (European) (for studies with mean age ≥ 40 + comorbidity > 40%)	PCR	23.68	10.85–36.51	13.00–39.97	100%
Qutub	Respiratory tract (severe patients)	Viral culture	47.5/20**	9.0–53.0		
Qutub	Respiratory tract (severe patients were not specified of excluded)	Viral culture	10/9**	8.0–13.0		
Qutub	Respiratory tract (immunocompromised patients)	Viral culture	54.36/20**	9.0–85.98		
Qutub	Respiratory tract (immunocompromised patients were not specified of excluded)	Viral culture	11.67/9**	8.2–13.3		
Rahmani	Not specific (immunocompetent individuals)	PCR	26.54	21.44–31.64	7.40–91.20	99.3%
Rahmani	Not specific (immunocompromised individuals)	PCR	36.28	21.93–50.63	15.90–132.00	94.2%
Xu	Respiratory tract (asymptomatic cases)	PCR	9.4±5.1***			
Xu	Gastrointestinal tract (asymptomatic cases)	PCR	16.8±9.8***			
Yan	Unrestricted (symptomatic cases)	PCR	19.7	17.2–22.7		99.34%
Yan	Unrestricted (asymptomatic cases)	PCR	10.9	8.3–14.3		98.89%
Yan	Unrestricted (severe patients)	PCR	24.3	18.9–31.1		91.88%
Yan	Unrestricted (nonsevere patients)	PCR	22.8	16.4–32.0		99.81%
Yan	Unrestricted (females)	PCR	19.4	9.5–39.4		93.93%
Yan	Unrestricted (males)	PCR	11.9	8.4–16.9		87.83%
Yan	Unrestricted (adults)	PCR	23.2	19.0–28.4		99.24%
Yan	Unrestricted (children)	PCR	9.9	8.1–12.2		85.74%
Yan	Unrestricted (with chronic diseases)	PCR	24.2	19.2–30.2		84.07%

Appendix 1—table 13 Continued on next page

Appendix 1—table 13 Continued

Study	Sampling site (subgroups)	Laboratory method	Pooled estimates (mean/median)	95% CI	Range	I ²
Yan	Unrestricted (without chronic diseases)	PCR	11.5	5.3–25.0		82.11%
Yan	Unrestricted (treated with corticosteroid)	PCR	28.3	25.6–31.2		0.00%
Yan	Unrestricted (treated without corticosteroid)	PCR	16.2	11.5–22.5		92.27%
Yan	Unrestricted (antiviral treatment)	PCR	17.6	13.4–22.2		98.99%
Yan	Unrestricted (mono-antiviral treatment)	PCR	21.2	15.3–29.2		90.04%
Yan	Unrestricted (multiantiviral treatment)	PCR	20.3	13.7–30.3		99.46%
Chen	Unrestricted (asymptomatic infections)	PCR or serum antibody	14.14*	11.25–17.04	11.00–17.25	

*Median estimate, **median/grouped median, ***this study analyzed by cases, and reported mean ± SD.

Appendix 1—table 14. Pooled estimates for duration of infectious period for COVID-19 in subgroups from two systematic reviews.

Study	Sampling site (subgroups)	Laboratory method	Pooled estimates	95% CI	Range	I ²
Rahmani	Replicant competent virus isolation	Viral culture	7.27	5.70–8.84	3.40–89.00	92.2%
Wang	Not specific	Not specific	6.25	5.09–7.51		
Rahmani	Replicant competent virus isolation (immunocompetent individuals)	Viral culture	6.33	4.92–7.75	3.00–13.00	92.4%
Rahmani	Replicant competent virus isolation (immunocompromised individuals)	Viral culture	29.50	12.46–46.53	13.80–89.00	84.8%

Appendix 1—table 15. Probability distributions of the incubation period and relative infectivity levels during the infectious period.

For the infectious period, day 0 corresponds to the symptom onset day. These two distributions are used to generate the infectiousness profile since infections.

Day	Incubation period		Infectious period	
	Mean = 5 days	Day	Max = 13 days	
1	0.058	–5	1.0	
2	0.11	–4	1.0	
3	0.14	–3	1.0	
4	0.16	–2	1.0	
5	0.15	–1	1.0	
6	0.13	0	1.0	
7	0.10	1	1.0	
8	0.068	2	1.0	
9	0.044	3	0.8	
10	0.026	4	0.6	
11	0.014	5	0.4	
12	0.0072	6	0.2	
13	0.0034	7	0.1	
14	0.0015	8		