Time-varying transmission heterogeneity of SARS and COVID-19 in Hong Kong

Dillon Adam
The University of Hong Kong
Katelyn Gostic
University of Chicago
Tim Tsang
University of Hong Kong https://orcid.org/0000-0001-5037-6776
Peng Wu (✉ pengwu@hku.hk)
The University of Hong Kong https://orcid.org/0000-0003-1157-9401
Wey Wen Lim
The University of Hong Kong
Amy Yeung
The University of Hong Kong
Jessica Wong
The University of Hong Kong
Eric Lau
University of Hong Kong https://orcid.org/0000-0002-6688-9637
Zhanwei Du
The University of Hong Kong
Dongxuan Chen
The University of Hong Kong
Lai-Ming Ho
The University of Hong Kong
Mario Martín-Sánchez
The University of Hong Kong
Simon Cauchemez
Institut Pasteur https://orcid.org/0000-0001-9186-4549
Sarah Cobey
University of Chicago https://orcid.org/0000-0001-5298-8979
Gabriel Leung
University of Hong Kong
Benjamin Cowling
University of Hong Kong https://orcid.org/0000-0002-6297-7154

Article

Keywords: coronavirus, COVID-19, SARS, superspreading, transmission, public health

Posted Date: March 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1407962/v1

Abstract

Transmission heterogeneity is a notable feature of the severe acute respiratory syndrome (SARS) and coronavirus disease 2019 (COVID-19) epidemics, though previous efforts to estimate how heterogeneity changes over time are limited. Using contact tracing data, we compared the epidemiology of SARS and COVID-19 infection in Hong Kong in 2003 and 2020-21 and estimated time-varying transmission heterogeneity (k_t) by fitting negative binomial models to offspring distributions generated across variable observation windows. k_t fluctuated over time for both COVID-19 and SARS on a continuous scale though SARS exhibited significantly greater ($p < 0.001$) heterogeneity compared to COVID-19 overall and in-time. For COVID-19, k_t declined over time and was significantly associated with increasingly stringent non-pharmaceutical interventions though similar evidence for SARS was inconclusive. Underdetection of sporadic COVID-19 cases led to a moderate overestimation of k_t, indicating COVID-19 heterogeneity of could be greater than observed. Time-varying or real-time estimates of transmission heterogeneity could become a critical indicator for epidemic intelligence in the future.

Main

Superspreading appears a distinct feature of coronavirus transmission during past and recent human outbreaks of novel coronaviruses including the ongoing coronavirus disease 2019 (COVID-19) pandemic (1–8). The effective reproductive number (*Re*), and time-varying effective reproduction number (Re_t), are essential epidemiological measures used to quantify rates of pathogen transmission and predict the progression of epidemics, however, these measures often omit heterogeneity in individual level transmission and superspreading effects. Typically, disease transmission is modelled according to a Poisson distribution where individual variance in transmission is equal to *Re*. However, this distribution is inappropriate when modelling datasets that feature superspreading because the variance greatly exceeds the expected mean. The negative binomial distribution is a convenient model to measure transmission heterogeneity when parameterized by the mean (μ), equivalent to *Re* and dispersion parameter, *k*, where the variance is equal to $\mu + \mu^2/k$ (9). For low values of k ($0 < k < 1$), the corresponding distribution is long tailed but concentrated around zero. Such overdispersion indicates the likelihood of superspreading events (9, 10) which can vastly alter epidemic dynamics and optimal intervention strategies (11). For example, highly overdispersed transmissions are often associated with large clusters of infection, thus interventions that aim to limit superspreading can have a greater marginal impact on *Re* (12–14). Unlike *Re* (and Re_t) however, *k* is traditionally interpreted as a fixed characteristic immutable to non-pharmaceutical interventions (NPIs), and few studies have investigated potential temporal changes in *k* (15, 16) and the relative effect of NPIs on transmission heterogeneity over time. Using detailed contact tracing data, we presented a novel approach to estimating transmission heterogeneity over time on a continuous scale, i.e., time-varying overdispersion (k_t) and other related statistics. We applied our approach to investigate and compare the epidemiology and temporal transmission heterogeneity with two coronavirus datasets from Hong Kong notable for superspreading: SARS in 2003, and the ongoing COVID-19 pandemic (1, 17).

Results

Comparative transmission dynamics of SARS and COVID-19 in Hong Kong

Figure 1 shows the epidemic curve of SARS in 2003 and COVID-19 in Hong Kong between Jan 23, 2020, and Apr 5, 2021. During the three epidemic waves of COVID-19 transmission, the majority of the local cases (n = 6,112/9,216, 66.34%) were not traceable to superspreading events (defined as greater than six secondary cases per infector case (1)). This was in contrast to SARS where the majority of the cases matched to contact tracing data (n = 848/1,295, 65.48%) were linked to superspreading events either as cases exposed directly to source of infection (primary SSE exposure cases) or subsequent generations traced to past events (other SSE associated cases). The mean overall estimate of Re_t for SARS was higher at 1.5 (range = 0.45, 5.8) compared to 1.1 (range = 0.37, 2.3) for COVID-19, though the range was notably wider for SARS (Fig. 2A). Consistent with the known kinetics of both pathogens, the mean serial interval for all resolved transmission pairs was 5.4 days for COVID-19 and 12.1 days for SARS. Temporally, the SARS epidemic was marked by a substantial rise in Re_t during the early phase of the outbreak in late February 2003 that preceded the peak in late March and coincided with a consistent decline and eventual plateau of Re_t throughout the final phase of the outbreak (Fig. 2A). This compared to COVID-19 where Re_t was marginally distributed around 1 but never peaked as high or declined as low as SARS during periods of local transmission (Fig. 2A).

The overall dispersion parameter for SARS in Hong Kong was $k = 0.04$ (Table 1). For COVID-19, we estimated a higher overall k of 0.20. From this we calculated the proportion of cases responsible for 80% of onwards transmissions (*Prop80*) as 4.2% for SARS in 2003, where 87.4% of cases did not transmit to anyone. In comparison, 14.2% of COVID-19 cases were responsible for 80% of onward infections in Hong Kong while fewer cases, 69.9%, did not transmit to anyone (Table 1, Supplementary Table 1).

Table 1
Overall estimates* of dispersion parameter (k) and proportion of cases responsible for infecting 80% (*Prop80*) and 0% (*Prop0*, i.e., no one) of cases for SARS and COVID-19 in Hong Kong.

	Dispersion k (95% CI)	*Prop80* (95% CI)	*Prop0* (95% CI)
SARS	0.04 (0.03–0.06)	4.2 (3.1–6.1)	87.4 (83.5–90.0)
COVID-19	0.20 (0.16–0.25)	14.2 (12.1–16.6)	69.9 (66.6–72.8)

*Overall estimates are a composite of all resolved and unresolved transmission pairs respectively. See methods for dataset construction criterion.

Time-varying transmission heterogeneity and dynamics of SARS & COVID-19 in Hong Kong

Figure 2B-C showed measures of time-varying transmission heterogeneity for COVID-19 and SARS in Hong Kong respectively by fitting negative binomial models to offspring distributions across variable sliding window periods (Supplementary Figure S1, See methods for details). For COVID-19, k_t and the time-varying *Prop80* ($Prop80_t$) fluctuated over time, however, during waves two and three both k_t and $Prop80_t$ significantly declined (Fig. 2B-C, Supplementary Table S2). Similar declines were inconclusive during the first COVID-19 wave seemingly due to the wider confidence intervals; a result of smaller sample sizes (Fig. 1). For SARS, we found that k_t and $Prop80_t$ appeared to significantly increase over time (Fig. 2B-C, Supplementary Table S2). Despite this contrasting temporal pattern, the final marginal density estimates of k_t for SARS remained significantly ($p < 0.001$) lower than for COVID-19 (Supplementary Table S3, Supplementary Figure S1), which was consistent with the overall estimates of *k* and *Prop80* in Table 1. This significant difference ($p < 0.001$) remained when accounting for differences in Re_t as shown by $Prop80_t$ finding that even in time, a smaller proportion of SARS cases were responsible for 80% of onward transmissions compared to COVID-19 (Fig. 2C, Supplementary Table S3, Supplementary Figure S1). For each epidemic (excluding COVID-19 wave one), given Re_t and k_t, we showed that the probability of epidemic extinction increased over time for both SARS and COVID-19 as expected given that each epidemic was locally eliminated (Fig. 2D, Supplementary Table S2,).

Temporal variations in k_t and $Prop80_t$ for COVID-19 appeared to partially correlate temporally with large superspreading events and clusters, particularly when case numbers were low and / or towards the declining phase of each epidemic (Fig. 1, Fig. 2B-C, Supplementary Figure S2). For example, during the third COVID-19 wave, a sharp decline in k_t and $Prop80_t$ coincided with a large gym cluster (n = 155) detected in mid-March 2021. Conversely, $Prop80_t$ and k_t were mostly insensitive to change following detection of the largest COVID-19 case cluster in Hong Kong during the initial epidemic rise in wave three. This cluster however was primarily associated with attendance across at least 21 dance and ballroom venues (n = 732), such that transmissions were unlikely to be caused by a single source case, while 48.9% (n = 355/732) of cases were secondary or tertiary cases etc. without direct exposure to the venues such as family members. Other apparent spikes in k_t and $Prop80_t$ could be seen throughout each COVID-19 wave that partially coincided with large superspreading events (Fig. 1, Fig. 2B-C, Supplementary Figure S1). For SARS, observed k_t and $Prop80_t$ were in contrast relatively monotonic following the initial introduction of a case from mainland China which triggered the first clusters linked to the Metropole Hotel (n = 17) and Prince of Wales Hospital (PWH, n = 143). Measures of heterogeneity appeared less sensitive to the occurrence of large clusters and superspreading events, most notably the Amoy Gardens estate outbreak (n = 329) and later hospital clusters (AHNH, n = 156) in late-Mar and early-Apr 2003 (Fig. 1)

We also identified a significant positive association between Re_t and measures of transmission heterogeneity (k_t and $Prop80_t$; $p = 0.003$ & < 0.001) for COVID-19 (Fig. 3A, Supplementary Table S4). This however was not the case for SARS where the relationship was inconclusive (Fig. 3, Supplementary Table S4). For COVID-19, increasing NPI stringency appeared to be associated with significant declines in k_t and $Prop80_t$, which was not seen for SARS as with Re_t (Fig. 3B, Supplementary Table S5). We did however

find a significant association between increasing NPI stringency and decreasing Re_t and increased probability of extinction as expected for both COVID-19 and SARS (Supplementary Figure S3, Supplementary Table S5).

Imperfect observation bias and underdetection

To investigate the potential effect of imperfect observation of COVID-19 cases on our estimates of k_t, we considered hypothetical worst-case scenarios involving under and over observation of chain-terminating cases in the data. To assess under-observation, we incorporated additional unobserved hypothetical cases as chain terminating singletons into the observed data. Here, the k_t estimated from the base case tended towards overestimation as expected (Fig. 4 left panel) i.e., the original k_t was higher than the k_t estimated from the sensitivity analysis on the hypothetical scenarios (Supplementary Figure S4). When testing a worst-case scenario where a constant 50% of singletons cases remained undetected, the k_t appeared to be overestimated with a median of 0.1 (90% CI: 0.05, 0.31) while the estimate was reduced to 0.03 (0.01, 0.08) under a more plausible scenario where the underdetection was 10% (Supplementary Table S6).

Next, to assess the upper boundary of potential over-observation, we excluded all observed singletons cases from the data while again testing a constant 10% – 50% underdetection rate as before. Here, k_t shifted upwards, as expected, compared to the estimate from the base case demonstrating the upper range for potential overestimation of k_t (Fig. 4 right panel). Excluding the hypothetical worst-case (50% underdetection), the marginal distribution of k_t for COVID-19 remained significantly higher than SARS (Supplementary Figure S5, Supplementary Table S7). Reassuringly, the absolute magnitude of bias was small, even in these worst-case-scenario analyses.

Discussion

This study presented for the first time a measurement of time-varying transmission heterogeneity (here k_t and $Prop80_t$) and compared two significant beta-coronavirus epidemics in Hong Kong: SARS in 2003 and the ongoing COVID-19 pandemic (Fig. 1 and Fig. 2). Few studies have demonstrated changing transmission heterogeneity between temporally distinct COVID-19 epidemic periods (15, 16), however we presented the variation in transmission heterogeneity on a continuous scale (daily) during periods of sustained local transmission with minimal to no international introductions which would otherwise confound local inference. We found that measures of transmission heterogeneity fluctuated over time and were partially correlated with large superspreading events, though measures for SARS exhibited significantly greater ($p < 0.001$) heterogeneity and were less variable over time compared to COVID-19 (Fig. 2, Supplementary Figure S2). Furthermore, k_t and $Prop80_t$ declined over the course of each epidemic wave. Counter to expectations, declines in k_t and $Prop80_t$ were significantly associated with increasingly stringent NPIs that largely targeted settings for potential superspreading events such as venue closures (Fig. 3D-E). A similar effect was observed in a recent study which saw overall estimates of k declined

after strict shelter-in-place measures were introduced (16). We found that k_t and $Prop80_t$ were also significantly associated with Re_t (Fig. 3A-B). Previous studies suggested this might be because population-wide NPIs, such as physical distancing and universal masking (that also reduced Re_t), effectively limited the number of random contacts in the community and thus increased the overall proportion of chain terminating cases, i.e., cases without onward infection, which is a hallmark feature of transmission heterogeneity (13, 18). Interpretation of k_t therefore is not simply proportional to the occurrence of superspreading events, though we did partially observe this, but individual transmissions overall. Thus, increasing measures of heterogeneity over time, including in response to NPIs, can indicate improving epidemic control as transmissions, including superspreading events, are prevented before they occur (13, 19). Evidence of the relative policy effect of these measures can be demonstrated by the repeat elimination of local COVID-19 infections prior to the wide availability of vaccines (20), unlike most countries which have seen widespread deaths and sustained circulation of COVID-19. Furthermore, estimation of k_t together with Re_t allows calculation of instantaneous probabilities of stochastic epidemic extinction, which like $Prop80_t$, is a more intuitive indicator for epidemic intelligence compared to k_t alone (Fig. 2.D, Supplementary Figure S6), and is undoubtedly improved over equivalent estimates when given fixed values of k.

This is also the first study to our knowledge to estimate overall overdispersion for SARS during the 2003 epidemic in Hong Kong, which was notable for the occurrence of large superspreading events. Our most conservative estimate of transmission heterogeneity (k = 0.05, 95% CI: 0.04, 0.07) however is lower than the only previous estimate for SARS (k = 0.16, 90% CI: 0.11, 0.64) using data from Singapore (9), demonstrating markedly greater transmission heterogeneity for SARS in Hong Kong (Table 1, Supplementary Table 1). Such difference between our results and others is perhaps indicative of a unique pathogen-population dynamic at the time. One major strength of our study therefore is the shared population setting, meaning comparisons between transmission heterogeneity for SARS and COVID-19 better accounts for potential confounders if comparing outbreaks between different populations not considering population changes between the epidemics. We have also refined previous estimates of COVID-19 transmission heterogeneity in Hong Kong by greatly increasing the number of transmission pairs available for analysis (n = 4,697 pairs minimum compared to n = 169 pairs for earlier estimates), producing additional evidence that favors greater levels of transmission heterogeneity for COVID-19 than previously found (Table 1, 12–17% vs 15–24% previously) (1). Our results are closer, but still higher than global estimates that showed < 10% of COVID-19 cases were responsible for 80% of transmissions (2, 3, 7, 21, 22). This consistent difference between overall estimates of heterogeneity between Hong Kong and elsewhere again could indicate a distinct pathogen-population dynamic in Hong Kong, though this might also be due to potential underdetection which we showed could lead to a moderate overestimation of k_t (Fig. 4).

Our study has some limitations. For COVID-19, no variants of concern (VOCs) sustained local transmission in Hong Kong during the study period, therefore our conclusions were based on originally dominant lineages and therefore might not be generalizable to VOCs, including those with altered

transmissibility e.g., Delta (B.1.617.2) (23, 24), Omicron (B.1.1.529) (25, 26) and other novel VOCs that may emerge in the future. This study also relied on contact tracing data largely generated through both active and passive surveillance in order to construct transmission pairs from which offspring distributions and measures of transmission heterogeneity were generated, and thus assumed all cases were observed. However as shown in our study, and consistent with past studies, imperfect observation and underdetection could lead to moderate overestimation of k (including k_t), thus underestimating the true degree of heterogeneity in transmission especially if singleton cases were over or underrepresented in the data. (Sporadic, chain-terminating cases are 0's in the offspring distribution; holding all other things equal, a higher number of observed 0's is a hallmark feature of greater overdispersion.) (18). If contact tracing fails to identify all valid epidemiological links, singleton cases are overrepresented. Alternatively, they could be underrepresented if singleton cases are harder to detect than large clusters which are more likely to observed (18). Observation biases are particularly problematic when the probability of observing epidemiological links between cases is low. However, in Hong Kong, COVID-19 observation probabilities are assumed to be relatively high given that each epidemic wave was effectively controlled throughout 2020 and 2021 via strict testing and trace, isolate and quarantine (20, 27), which makes us confident that relatively few infections are missing from our data, and the vast majority of observed COVID-19 cases were traced (Fig. 1). Regardless, under various worst case hypothetical scenarios, we showed that if 50% of COVID-19 cases in Hong Kong were undetected, the marginal density of k_t for COVID-19 was no longer significantly higher compared to SARS. However, when the same worst-case rate of underdetection was similarly applied to the overall estimates of k, heterogeneity remained significantly greater for SARS than for COVID-19 (k = 0.04, 95% CI: 0.03, 0.06 vs. k = 0.1, 95% CI: 0.08, 0.12 respectively).

Finally for SARS, it is likely that most cases were observed during the epidemic due to the relative severity, and thus the estimates of k_t for SARS were less likely to be overestimated due to underdetection compared with COVID-19. Instead, because 26.2% confirmed SARS cases could not be matched to available contact tracing data and were therefore excluded from all offspring-related analyses, which could result in bias in either direction.

Conclusion

Overall, we identified substantial individual heterogeneity in the transmissibility of both SARS and COVID-19, though SARS exhibited significantly greater ($p < 0.001$) heterogeneity compared to COVID-19 overall and in-time. Around 4% and 14% of cases were responsible for 80% of all onwards transmissions of SARS and COVID-19 in Hong Kong respectively, while 87 and 70% of cases did not transmit to anyone. We demonstrated for the first time estimates of temporal changing transmission heterogeneity on a continuous scale for both COVID-19 and SARS. Indicative of improving epidemic control of COVID-19, k_t significantly declined over the course of each epidemic wave and was significantly associated with increasingly stringent NPIs. Time-varying or real-time estimates of transmission heterogeneity could become a complementary indicator for epidemic intelligence during emerging epidemics and pandemics.

Declarations

COMPETING INTERESTS

BJC consults for AstraZeneca, Fosun Pharma, GSK, Moderna, Pfizer, Roche and Sanofi Pasteur. The authors report no other potential conflicts of interest.

AUTHOR CONTRIBUTIONS

All authors meet the ICMJE criteria for authorship. The study was conceived by DCA, GML and BJC. Data were collated from original sources by DCA, WWL, AY, JYW, EHYL, DC, LMH and MMS. Data were analyzed by DCA. DCA wrote the first draft of the manuscript, and all authors provided critical review and revision of the text and approved the final version.

ACKNOWLEDGMENTS

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 Scheme (Project No. C7123-20G) of the Research Grants Council of the Hong Kong SAR Government, and AIR@InnoHK administered by Innovation and Technology Commission. BJC is supported by a RGC Senior Research Fellow Scheme grant (HKU SRFS2021-7S03) from the Research Grants Council of the Hong Kong Special Administrative Region, China.

References

1. Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nature Medicine. 2020;26(11):1714-9.

2. Lemieux JE, Siddle KJ, Shaw BM, Loreth C, Schaffner SF, Gladden-Young A, et al. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. Science. 2021;371(6529):eabe3261.

3. Endo A, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res. 2020;5:67.

4. Cowling BJ, Park M, Fang VJ, Wu P, Leung GM, Wu JT. Preliminary epidemiological assessment of MERS-CoV outbreak in South Korea, May to June 2015. Euro Surveill. 2015;20(25):7-13.

5. Cho SY, Kang J-M, Ha YE, Park GE, Lee JY, Ko J-H, et al. MERS-CoV outbreak following a single patient exposure in an emergency room in South Korea: an epidemiological outbreak study. The Lancet. 2016;388(10048):994-1001.

6. Wang SX, Li Y, Sun B, Zhang S, Zhao W, Wei M, et al. The SARS outbreak in a general hospital in Tianjin, China–the case of super-spreader. Epidemiology & Infection. 2006;134(4):786-91.

7. Sun K, Wang W, Gao L, Wang Y, Luo K, Ren L, et al. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. Science. 2021;371(6526):eabe2424.

8. Du Z, Wang C, Liu C, Bai Y, Pei S, Adam DC, et al. Superspreading of SARS-CoV-2 infections: A Systematic Review and Meta-analysis. medRxiv. 2021:2021.12.09.21267507.

9. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438(7066):355-9.

10. Galvani AP, May RM. Dimensions of superspreading. Nature. 2005;438(7066):293-5.

11. Hébert-Dufresne L, Althouse BM, Scarpino SV, Allard A. Beyond R 0: heterogeneity in secondary infections and probabilistic epidemic forecasting. Journal of the Royal Society Interface. 2020;17(172):20200393.

12. Althaus CL. Ebola superspreading. The Lancet Infectious Diseases. 2015;15(5):507-8.

13. Sneppen K, Nielsen BF, Taylor RJ, Simonsen L. Overdispersion in COVID-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control. Proceedings of the National Academy of Sciences. 2021;118(14):e2016623118.

14. Frieden TR, Lee CT. Identifying and Interrupting Superspreading Events-Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis. 2020;26(6):1059-66.

15. Lim J-S, Noh E, Shim E, Ryu S. Temporal Changes in the Risk of Superspreading Events of Coronavirus Disease 2019. Open Forum Infectious Diseases. 2021;8(7).

16. Lau MSY, Grenfell B, Thomas M, Bryan M, Nelson K, Lopman B. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. Proceedings of the National Academy of Sciences. 2020;117(36):22430.

17. Yu ITS, Li Y, Wong TW, Tam W, Chan AT, Lee JHW, et al. Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus. New England Journal of Medicine. 2004;350(17):1731-9.

18. Blumberg S, Lloyd-Smith JO. Inference of R 0 and transmission heterogeneity from the size distribution of stuttering chains. PLoS computational biology. 2013;9(5):e1002993.

19. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health. 2020;8(4):e488-e96.

20. Gu H, Xie R, Adam DC, Tsui JL-H, Chu DK, Chang LD, et al. SARS-CoV-2 under an elimination strategy in Hong Kong. medRxiv. 2021.

21. Laxminarayan R, Wahl B, Dudala Shankar R, Gopal K, Mohan BC, Neelima S, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. Science. 2020;370(6517):691-7.

22. Miller D, Martin MA, Harel N, Tirosh O, Kustin T, Meir M, et al. Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. Nature Communications. 2020;11(1):5518.

23. Peacock TP, Sheppard CM, Brown JC, Goonawardane N, Zhou J, Whiteley M, et al. The SARS-CoV-2 variants associated with infections in India, B.1.617, show enhanced spike cleavage by furin. bioRxiv. 2021:2021.05.28.446163.

24. Liu Y, Liu J, Johnson BA, Xia H, Ku Z, Schindewolf C, et al. Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. bioRxiv. 2021:2021.08.12.456173.

25. Torjesen I. Covid-19: Omicron may be more transmissible than other variants and partly resistant to existing vaccines, scientists fear. BMJ. 2021;375:n2943.

26. Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature. 2022.

27. Wu P, Tsang TK, Wong JY, Ng TW, Ho F, Gao H, et al. Suppressing COVID-19 transmission in Hong Kong: an observational study of the first four months. 2020.

Online Methods

Characterization of clusters and transmission pairs

For COVID-19, we characterized case clusters and transmissions pairs confirmed between Jan 23, 2020, and Apr 5, 2021, using data provided by the Centre for Health Protection in Hong Kong, which was available for all 9,195 local cases detected during the study period. This period covered the epidemic waves of local transmission of COVID-19 in Hong Kong. Clusters linked to known imported index cases were excluded to limit inference of NPI effects to local measures only i.e., excluding travel related measures. Characterization, including the generation of transmission pairs, was performed as per our previous study with complete details of the construction process also available in the supplementary methods (1). Overall, we constructed two datasets each for COVID-19 and SARS sensitivity analyses. First for COVID-19, we identified 4,697 resolved transmission pairs (comprising 6,531 cases) where the specific infector and infectee could be identified (the primary dataset). Second, we additionally included cases whose infector was unknown but who could be traced to a common exposure source, i.e., as infectees linked directly to an exposure setting (hereafter denoted as unresolved transmission pairs & the sensitivity dataset), increasing the number of pairs to 6,196 pairs (comprising 7,640 cases). For SARS in 2003, contact tracing was performed and data provided by the Department of Health, and could be matched to demographic data for 1,293 (73.6%) of the 1,755 cases confirmed within Hong Kong. The 462 SARS cases with that could not be matched were excluded from all subsequent analyses. Generation of

transmission pairs was the same as for COVID-19, though we additionally reviewed publicly available reports to identify the confirmed index cases of large clusters where contact tracing data was ambiguous. As before, we constructed two datasets for sensitivity resulting in 472 transmission pairs (comprising 552 cases) for SARS where the specific infector and infectee was known, and 1,061 pairs (comprising 1,160 cases) when including infectees linked to common exposure settings.

Estimating transmission heterogeneity of SARS and COVID-19

Heterogeneity in transmission can be modelled according to a negative binomial distribution and corresponding dispersion parameter *k* fit to the individual numbers of secondary infections (offspring distribution), where the mean parameter, μ, is equivalent to the effective reproductive number (*Re*). Following the approach described by Lloyd Smith et al., (2) we generated offspring distributions for each of the primary and sensitivity datasets for SARS and COVID-19 by counting the number of secondary cases per primary case or exposure setting for all resolved and unresolved pairs. Cases without traceable onward transmission, including cases without a known source (unlinked singleton cases), were included in each distribution as chain terminating cases i.e., zero secondary cases, excluding those SARS cases without contact tracing data as before. Negative binomial distributions were fitted by maximum likelihood (ML) estimation to the primary and sensitivity offspring distributions independently using fitdistrplus v1.1-3 fixing *Re* = 1 for control interpretation of *k* between groups and datasets. Uncertainty around each point estimate of *k* was generated by parametric bootstrap resampling (n=1,000). Final overall estimates were a composite of the primary (upper bound) and sensitivity analyses (lower bound) calculated as the median and 95% quantiles of the combined bootstrapped distributions (See Supplementary Table S1 for independent results). To provide an intuitive comparison of overall transmission heterogeneity between SARS and COVID-19 on a linear scale, we calculated the proportion of individuals responsible for 80% of onward transmission (*Prop80*, by analogy to the 20/80 rule) as per the formula below given *Re* = 1 (3):

$$1 - 0.8 = \frac{1}{Re}\int_0^X \lfloor x \rfloor \mathrm{NegBin}\left(\lfloor x \rfloor; k, \frac{k}{Re + k}\right) dx$$

and which is eased by the following rearrangement

$$\frac{1}{Re}\int_0^X \lfloor x \rfloor \mathrm{NegBin}\left(\lfloor x \rfloor; k, \frac{k}{Re + k}\right) dx = \int_0^{X-1} \mathrm{NegBin}\left(\lfloor x \rfloor; k + 1, \frac{k}{Re + k}\right) dx$$

Estimating time-varying transmission heterogeneity

To estimate time-varying k, i.e., k_t, negative binomial distributions were fit by ML estimation to offspring distributions subset within dynamic sliding windows based on a minimum sample size and fixed window period forecast from each of the known infector's onset date, or confirmation date in the case of asymptomatic infection. When the infector was unknown, the infectees onset date or confirmation date when asymptomatic was used. For the dynamic windows we tested periods with a minimum of n = (10, 20, …, 50) paired cases. However, to avoid extending the dynamic windows beyond reasonable time periods when samples were low, windows were limited to one or two weeks regardless of the minimum sample size. Similarly, for the fixed windows, we subset offspring distributions into both one and two-week sliding windows. To simplify parameter estimation in time and control against biases in contact tracing data when case numbers were high, we estimated the median Re_t using EpiNow2 (4) for each of the dynamics and fixed sliding window periods which was then fixed as the input as parameter μ when estimating k_t. Uncertainty around each estimate was again generated by parametric bootstrap resampling. A final composite estimate of k_t was calculated by day as the median and 90% quantiles of the combined bootstrapped distribution of each of the variable window periods from both the primary and sensitivity datasets. We generated 90% CIs here instead of 95% for k_t because more extreme values of k were most difficult to estimate accurately with small sample sizes (5) as a result of sub-setting the observed data. This also allowed easier visualization of k through time and matched with the results as presented for SARS in Lloyd-Smith et al (2). With both Re_t and k_t known in time, we calculated the instantaneous probability of epidemic extinction given the probability generating function of a negative binomial offspring distribution (2) denoted by :

$$g(x) = \left(1 + \frac{Re_t}{k_t}(1 - x)\right)^{-k_t}$$

Where the probability that an epidemic has gone extinct by the n^{th} generation is $g^n(0).$ Complete details of the estimation procedure are available in the supplementary methods.

Temporal associations of changing transmission heterogeneity and NPIs

We conducted simple linear regression analyses to identify associations between changing transmission heterogeneity over time and temporally with Re_t and NPI stringency given the final estimates of k_t and $Prop80_t$. We performed 1000 independent regression analyses each randomly sampling within the

combined bootstrapped distribution of k_t and $Prop80_t$ n estimates equal to the number of days in each epidemic period with valid data. Significance was determined as the mean p-value of the 1000 estimates and adjusted by applying Bonferroni's correction for multiple comparisons. Because k_t exhibited a skewed distribution ($0 < k \rightarrow \infty$), values were transformed using the natural logarithm before linear regressions were conducted. The relative stringency of NPIs over time was calculated for COVID-19 and SARS based on the Oxford Government response tracker indices (level 1: <40; level 2 : 40–50; level 3: 50–60; level 4: 60–70; level 5: >70) (6). We compared the distributions of k_t between COVID-19 and SARS by Welch Two Sample t-test, performing 1000 independent comparisons sampled from the marginal bootstrapped distribution. Because k_t exhibits a skewed distribution ($0 < k \rightarrow \infty$), estimates were transformed using the natural logarithm before t-tests were conducted

Measuring bias due to incomplete observation and underdetection of COVID-19.

In order to investigate the magnitude and boundaries of possible bias due to the presence of an overburdened contact tracing system and asymptomatic infection, we considered worst-case scenarios that could cause us to underestimate or overestimate k_t. First, we could underestimate k_t if chain-terminating cases were overrepresented in our data (e.g., if epidemiological links between cases were frequently missed in contact tracing). To assess this worst-case underestimation bound, we repeated the main analysis excluding unlinked singleton cases observed in the community ("without singletons" analysis). Second, we could overestimate k_t if chain-terminating cases were systematically unobserved (e.g., if contact tracing made it easier to identify linked cases than singletons). We tested multiple worst-case hypothetical scenarios to investigate the overestimation boundary of k_t given extremes of potential observation bias by additionally incorporating unobserved singletons as chain terminating cases given assumed rates of underdetection (between 10 and 50% unobserved cases) as a proportion of the observed cases within each sliding window period prior to ML estimation of k. Together this conceptualizes the reasonable boundaries of k_t shown in Fig. 4. As before, we compared each the resulting hypothetical distributions of log-transformed k_t to the observed distribution for SARS by Welch Two Sample t-test, performing 1000 independent comparisons sampled from the marginal bootstrapped distribution.

DATA AVAILABILITY

All anonymized data is publicly available at https://github.com/dcadam/....

CODE AVAILABILITY

The code used for analysis is publicly available at https://github.com/dcadam/....

Figures

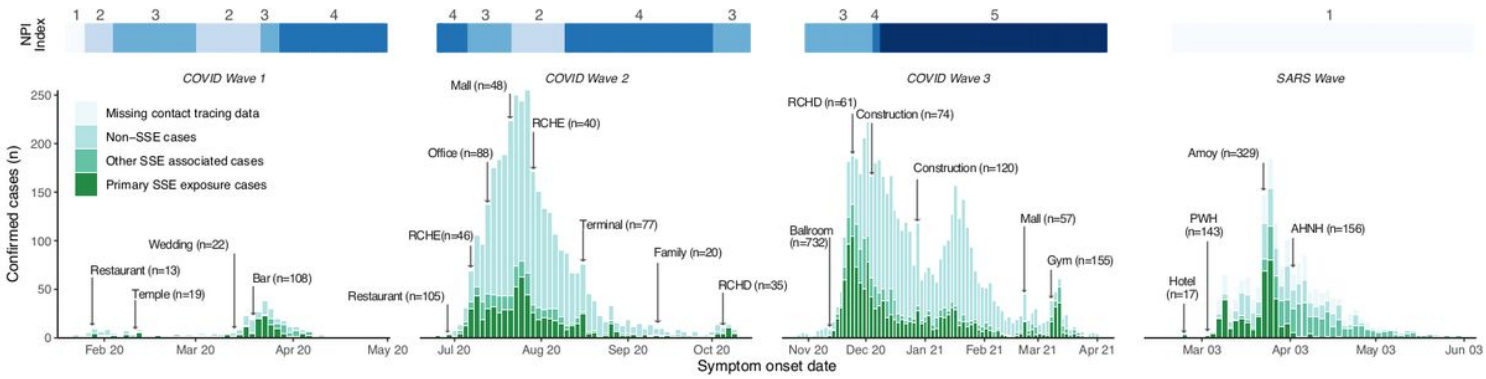

Figure 1

Epidemic curve of 9,237 locally acquired and confirmed COVID-19 cases, and 1,755 confirmed SARS cases in Hong Kong colored by source relative to known superspreading events. Arrows indicate the date of detection of the largest linked superspreading events and their primary setting including total number of linked cases overall including traceable onward infections.

Figure 2

A) Time-varying effective reproductive number of COVID-19 and SARS in Hong Kong estimated by EpiNow2. **B**) Estimate of time-varying overdispersion (k_t) with 90% confidence intervals shown as colored ranges around the median. **C)** Time-varying estimates of the proportion of cases infecting 80% of all other local cases ($Prop80_t$) with 90% confidence intervals shown as colored ranges calculated given Re_t and k_t. **D)** Time-varying estimates of the probability of stochastic epidemic extinction with 90% confidence intervals shown as colored ranges calculated given Re_t and k_t. For B-D, lines are plot given a linear regression model of the median points and the shaded range represents the standard error of the model.

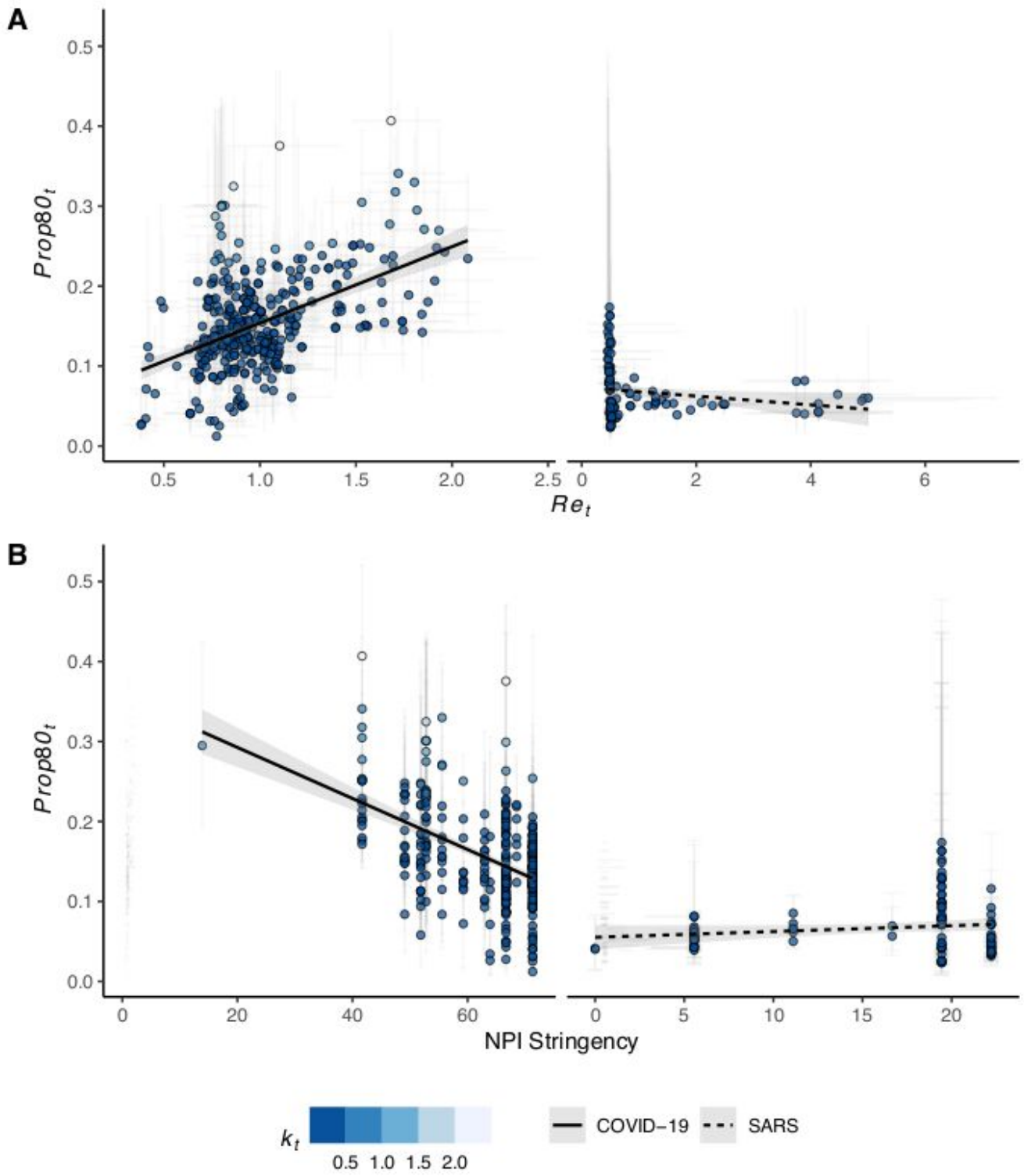

Figure 3

A) Relationship between the time-varying effective reproductive number (Re_t) and the proportion of cases responsible for 80% of transmission (Prop80t) for both COVID-19 (left) and SARS (right). (**B**) Relationship between Prop80t and increasing stringency of non-pharmaceutical interventions (NPI Stringency) for COVID-19 (left) and SARS (right). Points represent a sample of the median estimate with grey bars

reflecting the 90% CI generated by EpiNow2 and bootstrapped replicates for Re_t and Prop80t respectively. Both panels are colored by the estimated k_t.

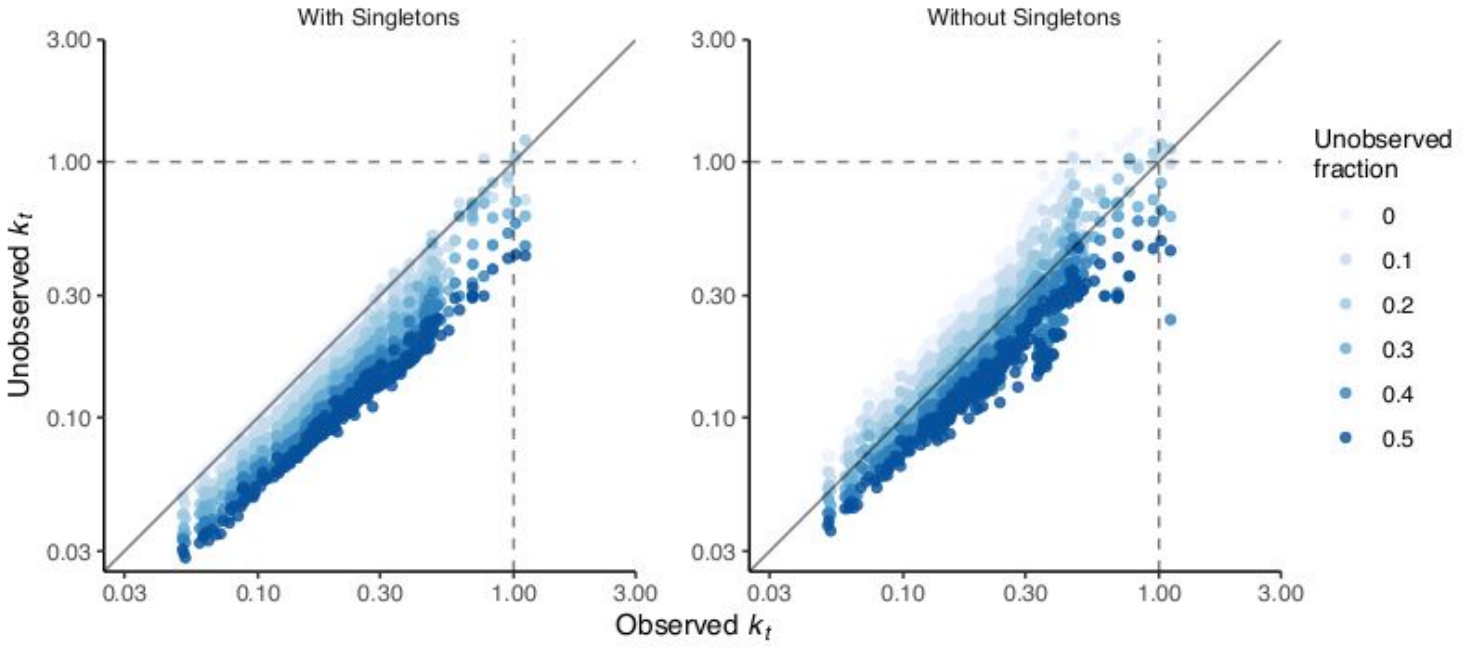

Figure 4

Direction and magnitude of bias between the observed k_t (Unobserved fraction = 0, with singletons) and estimates from hypothetical worst-case scenarios given various degrees of underdetection for COVID-19. Panel one compares the hypotheticals while including all observed unlinked singletons cases as chain terminating, which are excluded in panel two.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryfiles.docx