

1 **Effective real-time transmission estimations incorporating population viral load**
2 **distributions amid SARS-CoV-2 variants and pre-existing immunity**

3
4 **Potential Authors:**

5 Yu Meng¹, Yun Lin¹, Weijia Xiong¹, Eric H. Y. Lau^{1,2}, Faith Ho¹, Jessica Y. Wong¹, Peng Wu^{1,2},
6 Tim K. Tsang^{1,2}, Benjamin J. Cowling^{1,2}, Bingyi Yang^{1*}

7
8 **Affiliations:**

9 ¹ WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public
10 Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China.

11 ² Laboratory of Data Discovery for Health Limited, Hong Kong Science and Technology Park,
12 New Territories, Hong Kong, China

13
14 **Corresponding author:**

15 Bingyi Yang, School of Public Health, Li Ka Shing Faculty of Medicine, The University of
16 Hong Kong, 7 Sassoon Road, Pokfulam, Hong Kong Special Administrative Region, China

17 Email: yangby@hku.hk

18
19
20 Running head: Viral loads on SARS-CoV-2 transmission estimates

21 Word count (abstract): 267

22 Word count (main text): 3180

23 Keywords: SARS-CoV-2, viral loads, transmission, Omicron

24 **Abstract**

25 **Background:** Population-level viral load distribution, measured by cycle threshold (Ct), has
26 been demonstrated to enable real-time estimation of R_t for SARS-CoV-2 ancestral strain.
27 Generalisability of the framework under different circulating variants and pre-existing immunity
28 remains unclear.

29 **Aim:** This study aimed to examine the impact of evolving variants and population immunity on
30 the generalizability of Ct-based transmission estimation framework.

31 **Methods:** We obtained the first Ct record of local COVID-19 cases from July 2020 to January
32 2023 in Hong Kong. We modeled the association between daily viral load distribution and the
33 conventional estimates of R_t based on case count. We trained the model using data from wave 3
34 (i.e., ancestral strain with minimal population immunity) and predicted R_t for wave 5, 6 and 7
35 (i.e., omicron subvariants with > 70% vaccine coverage). Cross-validation was performed by
36 training on the other 4 waves. Stratification analysis by disease severity was conducted to
37 evaluate the impact of the changing severity profiles.

38 **Results:** Trained with the ancestral dominated wave 3, our model provided accurate estimation
39 of R_t , with the area under the ROC curve of 0.98 (95% confidence interval: 0.96, 1.00), 0.62 (95%
40 CI: 0.53, 0.70) and 0.80 (95% CI: 0.73, 0.88) for three omicron dominated waves 5 to 7,
41 respectively. Models trained on the other four waves also had high accuracy. Stratification
42 analysis suggested potential impact of case severity on model estimation, which coincided with
43 the fluctuation of sampling delay.

44 **Discussion:** Our findings suggested that incorporating population viral shedding can provide
45 accurate real-time estimation of transmission with evolving variants and population immunity.
46 Application of the model needs to account for sampling delay.

47 **Introduction**

48 Tracking community transmission in real-time is critical but suffers delays due to the
49 unavoidable right-censoring (i.e., incubation period and delay in case identification) in the
50 conventional methods.^{1,2} Previous work established a novel temporal association between
51 epidemic dynamics and population viral load distribution, measured by cycle threshold (Ct)
52 values from reverse transcription quantitative polymerase chain reaction (RT-qPCR).³ We have
53 since applied a simplified method to incorporate the temporal population Ct distribution into
54 real-time transmission estimation, measured by the effective reproductive number R_t .⁴ While
55 these studies have significantly advanced our understanding of the association between
56 population viral shedding and transmission, they were performed during early waves of the
57 COVID-19 pandemic, and did not account for the viral evolution and immunity derived from
58 natural infections and/or vaccinations.⁵ The generalizability of the identified association to
59 epidemics remained under-investigated, especially in the context of the emerging SARS-CoV-2
60 variants (e.g., Omicron) and increasing pre-existing immunity, which were found to be
61 associated with shorter duration in viral shedding clearance at individual level.⁶ Additionally,
62 large epidemics with exponential increase in cases could soon exceed testing and surveillance
63 capacity and may further prolong the delays between infection and being diagnosed due to the
64 constrained resources, making it more challenging to derive timely and reliable R_t estimates
65 using conventional incidence-based approaches.^{1,7}

66
67 Here, we examined the impact of the evolving SARS-CoV-2 variants and population immunity
68 on the application of population viral load distribution to estimate transmission, using data of
69 laboratory-confirmed COVID-19 cases from July 2020 to January 2023 in Hong Kong. During

70 this period, the dominant strains transitioned from ancestral strain to Omicron subvariants, and
71 the population shifted from predominantly native to possessing high pre-existing hybrid
72 immunity due to natural infection and vaccinations.⁸ Thus we can examine whether overall
73 association between population-level viral loads and transmission dynamics remained consistent
74 and validate the Ct-based R_t estimation method.

75

76 **Methods**

77 **Data source**

78 Viral loads of COVID-19 cases were measured as cycle threshold⁹ values (derived from SARS-
79 CoV-2 RT-qPCR assays targeting E gene) from upper respiratory tract samples, given that they
80 are inversely correlated (i.e. lower Ct values imply higher viral loads).^{10,11} We collected the
81 clinical, epidemiological and demographic data of each local cases from the Hospital Authority
82 (HA) and the Department of Health of the Government of Hong Kong during the observation
83 period, including their first recorded Ct values, date of sampling, clinical outcomes and date of
84 symptom onset. We also classified cases as mild-to-moderate, serious, critical and fatal
85 according to their clinical outcomes.^{8,12} In this study, we constrained analyses to two severity
86 groups, namely mild-to-moderate and severe groups (combining serious, critical and fatal cases
87 into a single group).

88

89 **Statistical Analysis**

90 ***R_t estimation based on case counts (incidence-based R_t)***

91 R_t estimation on COVID-19 should be based on infection time because of its pre-symptomatic
92 transmission^{1,7}. Thus, we applied robust incidence deconvolution estimator¹³ with delay from

93 infection to reporting, to reconstruct the epidemic curve by infection time. Then we estimate the
94 incidence-based R_t based on daily local cases numbers using Cori's method.⁷ In this framework,
95 the incidence-based R_t was the ratio of the number of new cases to the total infectiousness of
96 infective individuals, which was the convolution of the incubation period and the infectiousness
97 relative to onset.¹⁴ We conducted inference by a Markov chain Monte Carlo algorithm to
98 estimate R_t .^{15,16} More details about incidence-based R_t estimation was described elsewhere.^{1,4}

99

100 *Incorporating Ct distribution into R_t estimates (Ct-based R_t)*

101 Hong Kong experienced multiple epidemic waves between 1 January 2020 and 29 January 2023,
102 and it was dominated by local transmissions from wave 3.⁸ We analyzed the first record for
103 confirmed local COVID-19 cases (i.e., no travel outside Hong Kong during incubation) with
104 available Ct values. The whole observed period was split into three uninterrupted sub-periods,
105 i.e., 1 July to 31 August 2020 (wave 3), 1 November 2020 to 31 March 2021 (wave 4) and 1
106 January 2022 to 29 January 2023 (wave 5 and 6), to fit generalized additive model (GAM) to
107 characterize the population distribution of viral load separately (Appendix).

108

109 To assess the relationship between population-level distribution of viral loads and incidence-
110 based R_t , we first calculated the Spearman's rank correlation coefficient (ρ) between daily Ct
111 distribution (i.e., mean and skewness) and the natural log-transformed incidence-based R_t (Table
112 1). We fitted a linear regression model of the daily mean and skewness of Ct on log-transformed
113 incidence-based R_t (established from our previous study⁴), using training data from the third
114 wave (i.e., ancestral strain wave). We applied the fitted model to predict R_t using population-
115 level Ct distribution in the fourth to sixth waves (i.e., ancestral strain and Omicron variants

116 waves), respectively. We included a 31-day period in the training set (e.g., 19 July to 18 October
117 2020 for wave 3), consisting of 10 days before and 20 days after the day when local cases peaked
118 in that wave, as suggested by previous study.⁴ We determined the case peak by computing the 5-
119 day rolling average of confirmed local cases, to minimize the impact of sudden reporting
120 changes.

121
122 We evaluated the model prediction using the area under the receiver operating characteristic
123 curve (AUC) as the primary metric,^{17,18} fitted to the binary outcome of whether Ct-based
124 (predicted) and incidence-based (observed) R_t above 1. This threshold was chosen because an R_t
125 above 1 indicates a growing epidemic trend, while values below 1 indicate a decreasing trend,
126 which plays a crucial role in providing early warnings for public health.¹⁹ We also computed the
127 directional consistency as the proportion of days during the prediction period when the predicted
128 and observed R_t were either simultaneously below or exceeding 1.⁴

129
130 ***Cross-validation between epidemic waves***

131 To reflect the government's relaxation of entry restrictions and the introduction of new strains,
132 such as XBD, BF.7 and BQ.1.1, wave 6 was further split into waves 6 (23 May to 30 September
133 2022) and 7 (1 October 2022 to 29 January 2023).^{8,20} Then, we trained the model on data from
134 wave 4 to 7, separately, to evaluate the generalizability of our method. As described above, we
135 included a 31-day training period of wave 4 (24 November to 24 December 2020), wave 5 (21
136 February to 23 March 2022), wave 6 (23 August to 22 September 2022) and wave 7 (19
137 December 2022 to 18 January 2023), respectively. With each training set, the other waves as well
138 as wave 3 were used as test sets. We evaluated the model predictions using AUC and directional

139 consistency. We excluded from wave 5 the forecasts made between 1 January and 6 February
140 2022 due to the huge fluctuation of incidence-based R_t .

141

142 *Stratified analysis of two symptom severity groups*

143 To further validate the Ct-based framework, we assess the impact of the changing severity
144 profiles in confirmed COVID-19 cases. This involved characterizing the temporal changes of
145 first Ct distribution and delay of onset to sampling stratified by the retrospectively classified
146 clinical severity. In particular, the retrospective subsettings of severity were determined based on
147 their clinical outcomes that have already occurred, which may have been unknown during the
148 collection of the initial Ct values. Thus, we also characterized the distribution of the delays
149 between the initial records and clinical outcomes. Subsequently, we fitted the established model
150 from Ct data from either mild-to-moderate or severe group only in wave 3, 4, 5, 6, and 7,
151 respectively, and then used model and date from corresponding severity group to estimate the
152 scenarios in wave 5, 6 and 7.

153

154 All statistical analyses were conducted in R version 4.3.1 software (R Foundation for Statistical
155 Computing, Vienna, Austria).

156

157 **Results**

158 *COVID-19 waves in Hong Kong*

159 COVID-19 cases were detected from people with respiratory symptoms or high risk of exposures
160 (e.g., close contacts and occupational exposure) and confirmed with RT-qPCR between January
161 2020 and 6 February 2022,⁴ covering wave 3, 4 and early wave 5. Contact tracing were

162 suspended after 7 February 2022 (waves 5 and 6), and self-reported positive rapid antigen test
163 (RAT) were also recorded as cases after 26 February 2022.²¹ A decline in the incidence-based R_t
164 was observed throughout April and May 2022 before cases number increased again in June
165 partially due to the emergence of the Omicron BA.4/BA.5 which grew steadily and eventually
166 replaced BA.2 as the dominant variants, with BA.5 having an absolute advantage since later
167 August 2022.²² By later January 2023, sublineages BA.2 and BA.5 became the dominating
168 lineages in Hong Kong.²³ Two COVID-19 vaccines (CoronaVac [Sinovac] and BNT162b2
169 [BioNTech/Fosun Pharma/Pfizer]) were provided for free in Hong Kong government since
170 February 2021 (during and after wave 4).²⁴ As of January 2, 2022 (wave 5), approximately 66%
171 of the population had received at least one vaccine dose.²⁵

172
173 We included local cases from July 2020 to January 2023, covering wave 3 (1 July to 31 August
174 2020) and wave 4 (1 November 2020 to 31 March 2021) caused by the ancestral strain, while
175 wave 5 (1 January to 22 May 2022), wave 6 and 7 (23 May 2022 to 29 January 2023) were
176 caused by Omicron BA.2 and BA.4/BA.5, respectively (Figure).^{8,12}

177
178 In total 2,790,814 local RT-qPCR confirmed COVID-19 cases and self-reported RAT positives
179 were recorded through wave 3 and 7, with about 43%, 19% and 38% of the cases recorded in
180 wave 5, 6 and 7, respectively (Table S2). Ct values were available for 114,714 cases included,
181 with 95% (n=3,043), 96% (n=5,225), 4% (n=51,372), 4% (n=21,835), and 3% (n=33,239) in
182 waves 3 to 7, respectively.

183
184 ***Correlations between population-level Ct distribution and incidence-based R_t***

185 We examined the correlation between the temporal population-level distribution of Ct values
186 (measured by daily mean and skewness) and the local transmission dynamics (measured by the
187 incidence-based R_t). We observed consistent temporal associations between population Ct
188 distribution and incidence-based R_t in the Omicron variants-dominated waves 5 and 7, as seen
189 for the ancestral strain-dominated waves 3 and 4 (Figure). Higher incidence-based R_t were
190 observed with lower average Ct values (Spearman's correlation coefficient, $\rho = -0.69$, $P < 0.001$
191 for wave 5 and $\rho = -0.58$, $P < 0.001$ for wave 7) and as Ct skewed towards lower values ($\rho = 0.62$,
192 $P < 0.001$ for wave 5 and $\rho = 0.55$, $P < 0.001$ for wave 7) (Table 1), although such relationships
193 were not significant for wave 6 (Table 1).

194

195 ***Estimating Ct-based R_t for waves with changing dominant variant and population immunity***

196 We evaluated the performance of our approaches in later waves to examine the impact of
197 changing dominating SARS-CoV-2 variants (wave 4-7) and population immunity (wave 5-7) on
198 the Ct-based real-time COVID-19 transmission. We trained the model using the 31-day data
199 around peak of the ancestral strain-dominated wave 3 (as suggested previously⁴) and applied the
200 trained model to predict R_t using daily Ct distributions in waves 4 to 7 (i.e., testing periods),
201 separately. We found that the Ct-based method provided accurate real-time estimations of R_t
202 during testing periods (Table 2), with area under the receiver operating characteristic (ROC)
203 curve (AUC) of 0.68 (95% confidence interval (CI), 0.60, 0.75) (test set, wave 4), 0.98 (95% CI,
204 0.96, 1) (test set, wave 5), 0.62 (95% CI, 0.53, 0.70) (test set, wave 6) and 0.80 (95% CI, 0.73,
205 0.88) (test set, wave 7), respectively (Table 2).

206

207 We validated our methods by training the model using data from waves 4 to 7. The model
208 demonstrated high accuracy in predicted R_t , when using wave 3 and 5 as test sets (Table 2). For
209 example, AUC of predicted Ct-based R_t for Omicron variants-dominated wave 5 was 0.98 (95%
210 CI, 0.96, 1) (trained with wave 4), 0.99 (95% CI, 0.97, 1) (trained with wave 6) and 0.98 (95%
211 CI, 0.96, 1) (trained with wave 7). Similarly, retrospective R_t estimates for ancestral strain-
212 dominated wave 3 indicated robust accuracies with AUC of 0.92 (95% CI, 0.84, 1) (trained with
213 wave 4), 0.80 (95% CI, 0.70, 0.90) (trained with wave 5), 0.91 (95% CI, 0.83, 1) (trained with
214 wave 6) and 0.92 (95% CI, 0.85, 1) (trained with wave 7). However, despite accurate prediction
215 of wave 5, utilizing it as a training set led to suboptimal performance, as evidenced by the AUC
216 of predicted Ct-based R_t for wave 6 (0.53, 95% CI, 0.49, 0.57) and wave 7 (0.49, 95% CI, 0.46,
217 0.53). Of note, Ct-based estimates on wave 6 were generally weaker than other waves, with AUC
218 equal to 0.62 (95% CI, 0.53, 0.70) (trained with wave 3), 0.62 (95% CI, 0.53, 0.70) (trained with
219 wave 4), 0.53 (95% CI, 0.49, 0.57) (trained with wave 5), 0.66 (95% CI, 0.57, 0.75) (trained with
220 wave 7).

221

222 ***The impact of severity on the association between population-level Ct distribution and***
223 ***incidence-based R_t***

224 There were more severe (serious, critical, and fatal cases) cases recorded during Omicron waves
225 compared with ancestral strain waves. 40% cases were retrospectively classified as severe in
226 wave 5 and 7, compared with less than 25% in wave 3 and 4 (Table S3). As suggested, we
227 retrospectively assess Ct-based R_t estimates based on subsets of Ct values from cases with two
228 distinct degrees of severity, including mild-to-moderate group and severe group, to further
229 validate the generalizability of our model. A slight delay was found in the temporal distribution

230 of Ct values between mild-to-moderate and severe cases during wave 5 and 6 (Figure S2), which
231 coincided with the different delays of onset to sampling between mild-to-moderate and severe
232 group during February 2022 (wave 5), April 2022 (wave 5), and August 2022 (wave 6).

233
234 Results for the model trained in wave 3 suggested that using Ct values from mild-to-moderate or
235 severe only group yielded decreased AUCs (e.g., 0.52, 95% CI, 0.48, 0.56 for wave 5, using
236 severe cases only) compared to using data from all cases (e.g., 0.98, 95%CI, 0.96, 1 for wave 5,
237 using all cases), and the model trained in wave 4 show similar trend (Table S4). In contrast, for
238 the training period with increasing proportion of severe cases (Table S3), using Ct from one
239 severity group could yield higher AUCs compared to using Ct from all cases. For instance, the
240 AUC of Ct-based R_t for wave 6 was 0.63 (95% CI, 0.55, 0.71) (trained with mild-to-moderate
241 cases from wave 5), compared to 0.53 (95% CI, 0.49, 0.57) (trained with all cases from wave 5)
242 (Table S4).

243

244 **Discussion**

245 In this study, we demonstrated that the association between population viral load distribution and
246 epidemics, derived from ancestral strain with minimal population immunity, remained highly
247 informative during the subsequent SARS-CoV-2 epidemic waves, including those dominated by
248 the Omicron subvariants in populations with significant pre-existing immunity. Our findings also
249 suggested several circumstances in which the model predictions may be conservative, such as
250 amid fluctuations in sample representativeness or during plateaued local epidemics, thereby
251 providing insights into the broader applicability of the model.

252

253 Reduced viral shedding durations were observed for vaccinated individuals and those infected
254 with the Omicron variant; however, disparities in viral loads were minimal during the early
255 stages of infection.⁶ Therefore, our approach employing the initial Ct post-confirmation may be
256 minimally affected by variants and vaccinations, consistent with previous studies measuring the
257 transmission using various measurements of transmission (e.g., growth rate).^{5,26-28}

258
259 For surveillance based primarily on symptoms and contact tracing,⁴ Ct samples could be delays
260 in case identification or sample collection, resulting in an inaccurate reflection of the overall
261 viral load in the population. For example, we observed lower population Ct values for severe
262 cases in early wave 5 but higher values in late wave 5 and early wave 6, compared to mild-to-
263 moderate cases (Figure S2), coincided with observed disparities in delays from symptom onset to
264 sample collection (Figure S8). Our stratified analyses of sample severity profile also revealed a
265 complex impact of severity on the Ct-based estimation of R_t , highlighting the importance of
266 stable sample representativeness. During waves 3 and 4 in Hong Kong, when intensive
267 surveillance and contact-tracing adopted amid low virus circulation in the community, models
268 trained on all cases outperformed those trained only on mild-to-moderate cases, as they better
269 represented the full spectrum of case exposure time distribution (Figure S3-S7). Conversely,
270 when training the model using data that has significant increases in the proportion of severe
271 samples with delayed reporting, the model may inaccurately associate these changes in severity
272 profile and reporting to changes in transmission, leading to reduced model performance (Table
273 S4). This observation is consistent with the shorter delays observed between initial Ct reports
274 and clinical outcomes during Omicron waves compared to the delays observed during waves 3
275 and 4 (Figure S9). Future research on the impact surveillance delays and sample

276 representativeness on association between population viral shedding and transmission would
277 better inform the applicability of the Ct-based estimation framework.

278

279 The accuracy of Ct-based R_t estimates for wave 6 was not optimal, irrespective of whether the
280 model was trained with the data from the same wave or other waves, with deviations mainly
281 occurring in late May and early June 2022 (Figure S1). A possible reason could be the low and
282 relatively consistent community transmission, as evidenced by the incidence-based R_t stably
283 fluctuating around 1 and its Gini coefficient of 0.140 (95% CI, 0.125, 0.156) (Table S5).
284 Simultaneously, fewer samples available to obtain Ct values led to increased uncertainties
285 surrounding the R_t estimates, as suggested by our previous study.⁴ Whilst increased AUCs for
286 testing sets in wave 4 and 6 were observed when we excluded days with less than 30 or 60
287 records (Table S6). Besides, the training period included for wave 6 coincided with the presence
288 of multiple Omicron variants, including BA.2 and BA.4/BA.5. These variants were indicated to
289 exhibit distinct viral kinetics and infectiousness,^{29,30} however, limited genotyping data pose
290 challenges in evaluating the impact of the transition period when competing Omicron variant
291 coexisted on our model's performance. Future study on model predictions during co-circulation
292 or transition period are warranted for better understanding of our model applications under these
293 complex situations.

294

295 Our work has several limitations. Firstly, we were unable to further examine the effect of
296 vaccines and variants due to the limited individual data. Secondly, the reduction in case
297 ascertainment as non-pharmaceutical interventions were relaxed⁸ and the potential bias towards
298 reporting severe cases in the later epidemic waves may have affected the accuracy of R_t

299 estimation based on case counts. Further research is needed to address this issue and refine the
300 model accordingly. Additionally, it is worth noting that our model requiring less computation
301 efforts and crude metric AUC enables robust binomial estimation of R_t values exceeding 1 or
302 descending below 1. Whilst it may not afford absolute quantitative estimates, as was
303 demonstrated in the nowcast for wave 5 (Table S7 and Figure S1-C). Therefore, it is necessary to
304 validate the model in other populations to improve its generalizability.

305
306 Our study provides valuable insights into the potential of population-level Ct distribution as a
307 predictive tool for timely assessing R_t during waves characterized by variants dominating and
308 population immunity shifting. These findings suggest the potential generalizability of this
309 simplified framework across various settings and situations. It is important to exercise caution
310 when interpreting the results due to the fluctuation of sampling delay and severity proportion.
311 Further research is required to validate these findings and address the issue of estimating R_t
312 when daily records are limited and when community transmission were stable and consistent.

313
314 **Acknowledgements**

315 We thank the Department of Health and Hospital Authority of the Food and Health Bureau of the
316 Government of Hong Kong for providing the data for the analysis. This project was supported by
317 the Health and Medical Research Fund, Food and Health Bureau, Government of the Hong Kong
318 Special Administrative Region (grant no. 22210552, B.Y.), the Theme-based Research Scheme
319 (Project No. T11-705/21-N; B.J.C.) and the general research fund (Project No. 17100822; T.K.T.)
320 of the Research Grants Council of the Hong Kong SAR Government. We also thank Justin K.
321 Cheung and Chloe S. Chui for their help in managing and collecting the data.

322

323 **Author contributions**

324 All authors meet the ICMJE criteria for authorship. The study was conceived by B.Y. and B.J.C.
325 Y.M., Y.L., W.X., E.H.Y.L., F.H., J.Y.W., W.P. and T.K.T. prepared the data. B.Y., Y.M. and
326 Y.L. developed the model. Y.M. conducted the data analyses. Y.M., B.Y., T.K.T., and B.J.C.
327 interpreted the results. Y.M. wrote the first draft of the paper. All authors provided critical
328 review and revision of the text and approved the final version.

329

330 **Competing interests**

331 B.J.C. consults for AstraZeneca, Fosun Pharma, GSK, Haleon, Moderna, Roche, Sanofi Pasteur,
332 and Pfizer. The remaining authors declare no competing interests.

Reference

- 1 Tsang, T. K., Wu, P., Lau, E. H. Y. & Cowling, B. J. Accounting for Imported Cases in Estimating the Time-Varying Reproductive Number of Coronavirus Disease 2019 in Hong Kong. *J Infect Dis* **224**, 783-787 (2021). <https://doi.org/10.1093/infdis/jiab299>
- 2 Ho, F. *et al.* Accounting for the Potential of Overdispersion in Estimation of the Time-varying Reproduction Number. *Epidemiology* **34**, 201-205 (2023). <https://doi.org/10.1097/ede.0000000000001563>
- 3 Hay, J. A. *et al.* Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science* **373** (2021). <https://doi.org/10.1126/science.abh0635>
- 4 Lin, Y. *et al.* Incorporating temporal distribution of population-level viral load enables real-time estimation of COVID-19 transmission. *Nat Commun* **13**, 1155 (2022). <https://doi.org/10.1038/s41467-022-28812-9>
- 5 Sala, E. *et al.* Systematic Review on the Correlation Between SARS-CoV-2 Real-Time PCR Cycle Threshold Values and Epidemiological Trends. *Infect Dis Ther* **12**, 749-775 (2023). <https://doi.org/10.1007/s40121-023-00772-7>
- 6 Lin, Y. *et al.* Viral kinetics of SARS-CoV-2 following onset of COVID-19 in symptomatic patients infected with the ancestral strain and omicron BA.2 in Hong Kong: a retrospective observational study. *Lancet Microbe* **4**, e722-e731 (2023). [https://doi.org/10.1016/S2666-5247\(23\)00146-5](https://doi.org/10.1016/S2666-5247(23)00146-5)
- 7 Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol* **178**, 1505-1512 (2013). <https://doi.org/10.1093/aje/kwt133>
- 8 Yang, B. *et al.* Comparison of control and transmission of COVID-19 across epidemic waves in Hong Kong: an observational study. *Lancet Reg Health West Pac* **43**, 100969 (2024). <https://doi.org/10.1016/j.lanwpc.2023.100969>
- 9 Jones, T. C. *et al.* Estimating infectiousness throughout SARS-CoV-2 infection course. *Science (New York, N.Y.)* **373** (2021). <https://doi.org/10.1126/science.abi5273>
- 10 Tsui, E. L. H. *et al.* Development of a data-driven COVID-19 prognostication tool to inform triage and step-down care for hospitalised patients in Hong Kong: a population-based cohort study. *BMC Med Inform Decis Mak* **20**, 323 (2020). <https://doi.org/10.1186/s12911-020-01338-0>
- 11 Wong, C. K. H. *et al.* Viral burden rebound in hospitalised patients with COVID-19 receiving oral antivirals in Hong Kong: a population-wide retrospective cohort study. *Lancet Infect Dis* **23**, 683-695 (2023). [https://doi.org/10.1016/S1473-3099\(22\)00873-8](https://doi.org/10.1016/S1473-3099(22)00873-8)
- 12 Wong, J. Y. *et al.* Intrinsic and effective severity of COVID-19 cases infected with the ancestral strain and Omicron BA.2 variant in Hong Kong. *J Infect Dis* (2023). <https://doi.org/10.1093/infdis/jiad236>
- 13 Avadhanula, V. *et al.* Viral Load of Severe Acute Respiratory Syndrome Coronavirus 2 in Adults During the First and Second Wave of Coronavirus Disease 2019 Pandemic in Houston, Texas: The Potential of the Superspreader. *The Journal of Infectious Diseases* **223**, 1528-1537 (2021). <https://doi.org/10.1093/infdis/jiab097>
- 14 He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* **26**, 672-675 (2020). <https://doi.org/10.1038/s41591-020-0869-5>

- 15 Thompson, R. N. *et al.* Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* **29**, 100356 (2019). <https://doi.org/10.1016/j.epidem.2019.100356>
- 16 Salje, H. *et al.* Reconstruction of antibody dynamics and infection histories to evaluate dengue risk. *Nature* **557**, 719-723 (2018). <https://doi.org/10.1038/s41586-018-0157-4>
- 17 Janssens, A. & Martens, F. K. Reflection on modern methods: Revisiting the area under the ROC Curve. *Int J Epidemiol* **49**, 1397-1403 (2020). <https://doi.org/10.1093/ije/dyz274>
- 18 Dong, J. *et al.* Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. *Crit Care* **25**, 288 (2021). <https://doi.org/10.1186/s13054-021-03724-0>
- 19 Cowling, B. J. *et al.* Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *Lancet Public Health* **5**, e279-e288 (2020). [https://doi.org/10.1016/s2468-2667\(20\)30090-6](https://doi.org/10.1016/s2468-2667(20)30090-6)
- 20 The Government of the Hong Kong Special Administrative Region. COVID-19 & Flu express. Local Situation of COVID-19 Activity (as of October 21, 2022). , <https://www.chp.gov.hk/files/pdf/local_situation_covid19_en_20221021.pdf> (2022).
- 21 Real-time dashboard, <<https://covid19.sph.hku.hk/>> (2023).
- 22 The Government of the Hong Kong Special Administrative Region. Latest situation of COVID-19 (as of August 31, 2022). <https://www.chp.gov.hk/files/pdf/local_situation_covid19_en.pdf> (2022).
- 23 The Government of the Hong Kong Special Administrative Region. COVID-19 & Flu express. Local Situation of COVID-19 Activity (as of February 8, 2023). , <<https://www.chp.gov.hk/en/resources/29/100148.html>> (2023).
- 24 McMenamin, M. E. *et al.* Vaccine effectiveness of one, two, and three doses of BNT162b2 and CoronaVac against COVID-19 in Hong Kong: a population-based observational study. *Lancet Infect Dis* **22**, 1435-1443 (2022). [https://doi.org/10.1016/s1473-3099\(22\)00345-0](https://doi.org/10.1016/s1473-3099(22)00345-0)
- 25 The Government of the Hong Kong Special Administrative Region. Hong Kong Vaccination Dashboard. 2022. (accessed January 2, 2022). <<https://www.covidvaccine.gov.hk/en/>> (2022).
- 26 Tso, C. F., Garikipati, A., Green-Saxena, A., Mao, Q. & Das, R. Correlation of Population SARS-CoV-2 Cycle Threshold Values to Local Disease Dynamics: Exploratory Observational Study. *JMIR Public Health Surveill* **7**, e28265 (2021). <https://doi.org/10.2196/28265>
- 27 Stevens, R., Pratama, R., Naing, Z. & Condylis, A. Analysis of SARS-CoV-2 real-time PCR test CT values across a population may afford useful information to assist public health efforts and add refinement to epidemiological models. *Pathology* **54**, 800-802 (2022). <https://doi.org/10.1016/j.pathol.2022.07.003>
- 28 Mishra, B. *et al.* High proportion of low cycle threshold value as an early indicator of COVID-19 surge. *J Med Virol* **94**, 240-245 (2022). <https://doi.org/10.1002/jmv.27307>
- 29 Wang, Q. *et al.* Antibody evasion by SARS-CoV-2 Omicron subvariants BA.2.12.1, BA.4 and BA.5. *Nature* **608**, 603-608 (2022). <https://doi.org/10.1038/s41586-022-05053-w>

- 30 Hay, J. A., Kennedy-Shaffer, L. & Mina, M. J. Viral loads observed under competing strain dynamics. *medRxiv*, 2021.2007.2027.21261224 (2021). <https://doi.org/10.1101/2021.07.27.21261224>

Figure. Temporal distribution of incidence-based R_t (estimated by cases count), and population-level Ct values (measured by daily mean and skewness). **a.** Locally confirmed COVID-19 cases with available Ct values by date of reporting and incidence-based R_t estimated by cases count. Black bars indicate daily case counts. Lines and shaded areas indicate the means and 95% credible intervals (CrIs) for incidence-based R_t over the entire observed period. **b.** Temporal distribution of population-level Ct values and main dominant strains. Black bars indicate the number of daily collected samples. Brown lines and shaded areas indicate the average and 95% confidence intervals (CIs) of Ct values estimated by a generalized additive model (GAM) during the third and fourth waves which were dominated by ancestral strain. Orange lines and shaded areas correspond to the fifth and sixth waves which were dominated by Omicron variants. **c.** Temporal distribution of Ct skewness. Dots and vertical lines represent the mean and 95% CIs of daily Ct skewness. **d, e.** Correlations between the incidence-based R_t and Ct mean (panel d) or skewness (panel e) during the fifth and sixth waves. Box plots indicate the interquartile ranges (IQR) and medians of the incidence-based R_t under various intervals of daily Ct mean (panel d) and skewness (panel e).

Table 1. Spearman's correlation coefficients (ρ) between Ct and the natural log-transformed incidence-based R_t .

	Wave 3 (Jul - Aug 2020)		Wave 4 (Nov 2020 - Mar 2021)		Wave 5 (Jan 2022 - May 2022)		Wave 6 (May 2022 - Sep 2022)		Wave 7 (Oct 2022 - Jan 2023)	
	ρ	<i>P</i> -value [^]	ρ	<i>P</i> -value [^]	ρ	<i>P</i> -value [^]	ρ	<i>P</i> -value [^]	ρ	<i>P</i> -value [^]
Ct mean	-0.79	<0.001	-0.51	<0.001	-0.69	<0.001	-0.14	0.105	-0.58	<0.001
Ct skewness	0.8	<0.001	0.27	0.001	0.62	<0.001	0.11	0.221	0.55	<0.001

[^]Two-sided P-values that were rounded to 3 decimal places.

Table 2. Model performance using different training periods to estimate Ct-based R_t in the other four waves.

	Training period: wave 3 [*] (19 Jul-18 Aug 2020)		Training period: wave 4 (24 Nov - 24 Dec 2020)		Training period: wave 5 (21 Feb -23 Mar 2022)		Training period: wave 6 (23 Aug-22 Sep 2022)		Training period: wave 7 (19 Dec 2022-18 Jan 2023)	
	AUC ^a	Consistency ^b	AUC ^a	Consistency ^b	AUC ^a	Consistency ^b	AUC ^a	Consistency ^b	AUC ^a	Consistency ^b
Wave 3	0.94 (0.86, 1.00)	94.7%	0.92 (0.84, 1.00)	94.7%	0.80 (0.70, 0.90)	77.2%	0.91 (0.83, 1.00)	93.0%	0.92 (0.85, 1.00)	93.0%
Wave 4	0.68 (0.60, 0.75)	71.9%	0.69 (0.61, 0.76)	73.3%	0.67 (0.59, 0.75)	66.4%	0.70 (0.62, 0.77)	74.7%	0.71 (0.64, 0.79)	74.7%
Wave 5	0.98 (0.96, 1.00)	97.1%	0.98 (0.96, 1.00)	97.1%	0.96 (0.92, 1.00)	96.2%	0.99 (0.97, 1.00)	98.1%	0.98 (0.96, 1.00)	97.1%
Wave 6	0.62 (0.53, 0.70)	67.9%	0.62 (0.53, 0.70)	67.9%	0.53 (0.49, 0.57)	35.9%	0.66 (0.58, 0.74)	59.5%	0.66 (0.57, 0.75)	68.7%
Wave 7	0.80 (0.73, 0.88)	80.2%	0.81 (0.73, 0.88)	81.0%	0.49 (0.46, 0.53)	37.2%	0.53 (0.50, 0.56)	41.3%	0.67 (0.59, 0.75)	62.8%
Overall^c	0.78 (0.74, 0.82)	78.1%	0.83 (0.79, 0.86)	82.9%	0.53 (0.49, 0.57)	51.2%	0.69 (0.65, 0.73)	73.4%	0.80 (0.76, 0.84)	80.6%

Incidence-based R_t was natural log-transformed. ^{*} The main model used to estimate Ct-based R_t

a AUC: area under the receiver operating characteristic curve (ROC)

b Directional consistency.

c Overall: combined all test sets into one to calculate corresponding AUC and directional consistency

