The Journal of Infectious Diseases

MAJOR ARTICLE

Effective Real-time Transmission Estimations Incorporating Population Viral Load Distributions Amid SARS-CoV-2 Variants and Preexisting Immunity

Yu Meng,[1] Yun Lin,[1] Weijia Xiong,[1,©] Eric H. Y. Lau,[1,2,©] Faith Ho,[1] Jessica Y. Wong,[1] Peng Wu,[1,2,©] Tim K. Tsang,[1,2] Benjamin J. Cowling,[1,2,©] and Bingyi Yang[1,©]

[1]WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong; and [2]Laboratory of Data Discovery for Health Limited, Hong Kong Science and Technology Park, China

Background. Population-level cycle threshold (Ct) distribution allows for R_t estimation for SARS-CoV-2 ancestral strain, however, its generalizability under different circulating variants and preexisting immunity remains unclear.

Methods. We obtained the first Ct record of local COVID-19 cases from July 2020 to January 2023 in Hong Kong. The log-linear regression model, fitting on daily Ct mean and skewness to R_t estimated by case count, was trained with data from ancestral-dominated wave (minimal population immunity), and we predicted the R_t for Omicron waves (>70% vaccine coverage). Cross-validation was performed by training on other waves. Stratification analysis was conducted to retrospectively evaluate the impact of the changing severity profiles.

Results. Model trained with the ancestral-dominated wave accurately estimated whether R_t was >1, with areas under the receiver operating characteristic curve of 0.98 (95% CI, 0.96–1.00), 0.62 (95% CI, 0.53–0.70), and 0.80 (95% CI, 0.73–0.88) for Omicron-dominated waves, respectively. Models trained on other waves also had discriminative performance. Stratification analysis suggested the potential impact of case severity on model estimation, which coincided with sampling delay.

Conclusions. Incorporating population viral shedding can provide timely and accurate transmission estimation with evolving variants and population immunity, though model application should consider sampling delay.

Keywords. reproduction number; SARS-CoV-2; transmission; viral loads; omicron variants.

Tracking community transmission in real time is critical but suffers delays due to the unavoidable right censoring (ie, incubation period and delay in case identification) in the conventional methods [1, 2]. Previous work established a novel temporal association between epidemic dynamics and average population cycle threshold (Ct) values measured from reverse transcription quantitative polymerase chain reaction (RT-qPCR) among positive specimens of SARS-CoV-2 [3]. At the individual level, viral loads peak around symptom onset and decrease over time, resulting in higher Ct values as the intervals between infection and testing become longer [4, 5]. Consequently, a growing epidemic dominated by recent infections will have higher viral loads and lower Ct values, while a declining epidemic dominated by older infections will have lower viral loads and higher Ct values [3]. Using this rationale, we have since applied a simplified method to incorporate the temporal population Ct distribution into real-time transmission estimation, measured by the effective reproductive number R_t [4].

While these studies have significantly advanced our understanding of the association between population viral shedding and transmission, they were performed during early waves of the COVID-19 pandemic [6]. The generalizability of the identified association to epidemics remained underinvestigated, especially in the context of the emerging SARS-CoV-2 variants (eg, Omicron) and increasing preexisting immunity, which were found to be associated with shorter duration in viral shedding clearance at the individual level [7]. Additionally, large epidemics with an exponential increase in cases could soon exceed testing and surveillance capacity and may further prolong the delays between infection and diagnosis due to the constrained resources, making it more challenging to derive timely and reliable R_t estimates with conventional incidence-based approaches [1, 8].

Here, we examined the impact of the evolving SARS-CoV-2 variants and population immunity on the application of population viral load distribution to estimate transmission, using data of laboratory-confirmed COVID-19 cases from July 2020 to

Received 08 July 2024; editorial decision 20 November 2024; accepted 25 November 2024; published online 27 November 2024

Presented in part: Poster presentation at the 9th Epidemics International Conference on Infectious Disease Dynamics 2023, Bologna, Italy, 28 November–1 December 2023.

Correspondence: Bingyi Yang, PhD, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 7 Sassoon Road, Pokfulam, Hong Kong Special Administrative Region, China (byyang@connect.hku.hk).

The Journal of Infectious Diseases® 2025;231:684–91

https://doi.org/10.1093/infdis/jiae592

January 2023 in Hong Kong. During this period, dominant strains transitioned from the ancestral strain to Omicron subvariants, and the population shifted from predominantly native to possessing high preexisting hybrid immunity due to natural infection and vaccinations [9]. Additionally, we conducted secondary analyses using Ct data from specific severity groups to predict transmission across the entire population, thereby exploring the impact of changing severity profiles on the model generalizability. Thus, we can examine whether overall association between population-level viral loads and transmission dynamics remained consistent and validate the Ct-based R_t estimation method.

METHODS

Data Source

Viral loads of COVID-19 cases were proxied by Ct [5] values (derived from SARS-CoV-2 RT-qPCR assays targeting the E gene) from upper respiratory tract samples, given that they are inversely correlated (ie, lower Ct values imply higher viral loads) [10, 11]. We collected the clinical, epidemiologic, and demographic data of each local case from the Hospital Authority and the Department of Health of the Government of Hong Kong during the observation period, including the first recorded Ct value, date of sampling, clinical outcomes, and date of symptom onset. The confirmed cases were classified as mild to moderate, serious, critical, and fatal according to their clinical outcomes, including the assessment of oxygen desaturation, medication/procedure use, and health care utilization data [9, 12]. This facilitated a secondary analysis to further explain the generalizability of a Ct-based framework.

Statistical Analysis

R_t Estimation Based on Case Counts: Incidence-Based R_t

We applied a robust incidence deconvolution estimator [13] with delay from infection to reporting to reconstruct the epidemic curve by infection time. Then we estimated the incidence-based R_t according to daily local case numbers using the method of Cori et al [8]. In this framework, the incidence-based R_t was the ratio of the number of new cases to the total infectiousness of individuals who were infective, which was the convolution of the incubation period and the infectiousness relative to onset [14]. We conducted inference by a Markov chain Monte Carlo algorithm to estimate R_t [15, 16]. More details about incidence-based R_t estimation are described elsewhere [1, 4].

Incorporating Ct Distribution Into R_t Estimates: Ct-Based R_t

Hong Kong experienced multiple epidemic waves between 1 January 2020 and 29 January 2023 [9]. We analyzed the first records for confirmed local COVID-19 cases (ie, no travel outside Hong Kong during incubation) with available Ct values. The whole observed period was split into 3 uninterrupted subperiods—1 July to 31 August 2020 (wave 3), 1 November 2020 to 31 March 2021 (wave 4), and 1 January 2022 to 29 January 2023 (wave 5–7)—to fit a generalized additive model to characterize the population distribution of viral load separately (supplementary appendix).

Of note, wave 6 (23 May–30 September 2022) and wave 7 (1 October 2022–29 January 2023) were separated to reflect the government's relaxation of entry restrictions and the introduction of new strains, such as XBD, BF.7, and BQ.1.1 [9, 17]. To assess the relationship between population-level distribution of viral loads and incidence-based R_t, we first calculated the Spearman's rank correlation coefficient (ρ) between daily Ct distribution (ie, mean and skewness) and the natural log-transformed incidence-based R_t for each of 5 waves (Table 1). This study aims to validate the generalizability of association between population Ct distribution and transmission epidemics and the simplified Ct-based model [4] in the context of SARS-CoV-2 variants and preexisting immunity. In our previous work using training data from wave 3 [4], the linear regression model of daily Ct mean and skewness on log-transformed incidence-based R_t performed the best [4]. We then applied the fitted model to predict R_t in waves 4 to 7. We included a 31-day period in the training set (eg, 19 July–18 August 2020 for wave 3), consisting of 10 days before and 20 days after the day when local cases peaked in that wave, as suggested by previous study [4]. We determined the case peak by computing the 5-day rolling average of confirmed local cases to minimize the impact of sudden reporting changes.

We evaluated model prediction using the area under the receiver operating characteristic curve (AUC) as the primary metric [18, 19], fitted to the binary outcome of whether the Ct-based R_t (predicted) or incidence-based R_t (observed) was >1. This threshold was chosen because an R_t >1 indicates a growing epidemic trend, while values <1 indicate a decreasing trend, which plays a crucial role in providing early warnings for public health [20]. An AUC <0.50 is considered nondiscriminative; between 0.50 and 0.69, discriminative; between 0.70 and 0.79, acceptable; between 0.80 and 0.89, excellent; and ≥0.90, outstanding [21]. We also computed mean absolute percentage error (MAPE) to reflect the numerical accuracy.

Cross-validation Between Epidemic Waves

We trained the model on data from waves 4 to 7, separately, to further evaluate the generalizability of our method. As described previously, we included a 31-day training period of wave 4 (24 November–24 December 2020), wave 5 (21 February–23 March 2022), wave 6 (23 August–22 September 2022) and wave 7 (19 December 2022–18 January 2023). With each training set, all 5 complete waves are automatically set as the test sets, including the wave where the training set is located. We evaluated the model predictions using AUC and MAPE. We excluded from wave 5 the forecasts made between 1 January and 6 February 2022 due to the huge fluctuation of incidence-based R_t.

Table 1. Spearman 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	
Ct	ρ	*P* Value	ρ	*P* Value	ρ	*P* Value	ρ	*P* Value	ρ	*P* Value
Mean	−0.79	<.001	−0.51	<.001	−0.69	<.001	−0.14	.105	−0.58	<.001
Skewness	0.80	<.001	0.27	.001	0.62	<.001	0.11	.221	0.55	<.001

Two-sided *P* values were rounded to 3 decimal places.

Abbreviations: Ct, cycle threshold; R_t, effective reproductive number.

Stratified Analysis of 2 Symptom Severity Groups

To further validate the Ct-based framework and assess the potential for prediction via targeted training on a subset of cases, we conducted an exploratory analysis to investigate the impact of the changing severity profiles on model performance. This involved characterizing the temporal changes of first Ct distribution and delay of onset to sampling stratified by the retrospectively classified clinical severity. We constrained analyses to 2 groups—namely, mild to moderate and severe—combining serious, critical, and fatal cases into a single group for the latter (supplementary appendix). In particular, the retrospective subsettings of severity were determined by their clinical outcomes that have already occurred, which may have been unknown during the collection of the initial Ct values. Thus, we also characterized the distribution of the delays between the initial records and clinical outcomes. Subsequently, we fitted the established model from Ct data from either the mild to moderate or severe group only in waves 3 to 7 and then used model and date from the corresponding severity group to estimate the scenarios in waves 5 to 7.

All statistical analyses were conducted in R version 4.3.1 (R Foundation for Statistical Computing).

RESULTS

COVID-19 Waves in Hong Kong

COVID-19 cases were detected from people with respiratory symptoms or high risk of exposures (eg, close contacts and occupational exposure) and confirmed with RT-qPCR between January 2020 and 6 February 2022 [4], covering waves 3 and 4 and the early wave 5. Contact tracing was suspended after 7 February 2022 (waves 5–7), and self-reported positive rapid antigen test (RAT) results were recorded as cases after 26 February 2022 [22]. A decline in the incidence-based R_t was observed throughout April and May 2022 before case numbers increased again in June, partially due to the emergence of Omicron BA.4/BA.5, which grew steadily and eventually replaced BA.2 as the dominant variants, with BA.5 having an absolute advantage since later August 2022 [23]. By later January 2023, sublineages BA.2 and BA.5 became the dominating lineages in Hong Kong [24]. Two COVID-19 vaccines (CoronaVac [Sinovac]; BNT162b2 [BioNTech/Fosun Pharma/Pfizer]) were provided in Hong Kong since February 2021, during and after wave 4 [25]. As of 2 January 2022 (wave 5), approximately 66% of the population had received at least 1 vaccine dose [26].

We included local cases from July 2020 to January 2023, covering wave 3 (1 July–31 August 2020) and wave 4 (1 November 2020–31 March 2021) caused by the ancestral strain, while wave 5 (1 January–22 May 2022) and waves 6 and 7 (23 May 2022–29 January 2023) were caused by Omicron BA.2 and BA.4/BA.5, respectively (Figure 1) [9, 12].

In total 2 790 814 local RT-qPCR–confirmed COVID-19 cases and self-reported RAT-positive results were recorded through waves 3 and 7 (Supplementary Table 2). Among these local confirmed cases, 100% (n = 3217), 100% (n = 5426), 62% (n = 746 829), 32% (n = 170 413), and 21% (n = 219 095) were confirmed by RT-qPCR in waves 3 to 7, respectively. However, Ct values were available for 114 714 cases included, with 95% (n = 3043), 96% (n = 5225), 7% (n = 51 372), 13% (n = 21 835), and 15% (n = 33 239) for RT-qPCR-positive cases in waves 3 to 7.

Correlations Between Population-Level Ct Distribution and Incidence-Based R_t

We observed consistent temporal associations between population Ct distribution and incidence-based R_t in the Omicron variants–dominated waves 5 and 7, as seen for the ancestral strain–dominated waves 3 and 4 (Figures 1 and 2). Overall, waves 3 and 4 were marked by low community virus circulation, with initially elevated incidence-based R_t that subsequently plateaued. In contrast, wave 5 exhibited the highest incidence early on, but transmission rapidly declined after February 2022. Waves 6 and 7 suggested a more stable and consistent transmission pattern throughout the observation period. Higher incidence-based R_t was observed with lower average Ct values (Spearman's rank correlation coefficient, $\rho = -0.69$ [$P < .001$] for wave 5; $\rho = -0.58$ [$P < .001$] for wave 7) and as Ct skewed toward lower values ($\rho = 0.62$ [$P < .001$] for wave 5; $\rho = 0.55$ [$P < 0.001$] for wave 7), although such relationships were not significant for wave 6 (Table 1).

Estimating Ct-Based R_t for Waves With Changing Dominant Variant and Population Immunity

The model was trained on 31-day data around the peak of the ancestral strain–dominated wave 3, as suggested previously [4]. It was accurate to predict whether R_t is >1 in the subsequent

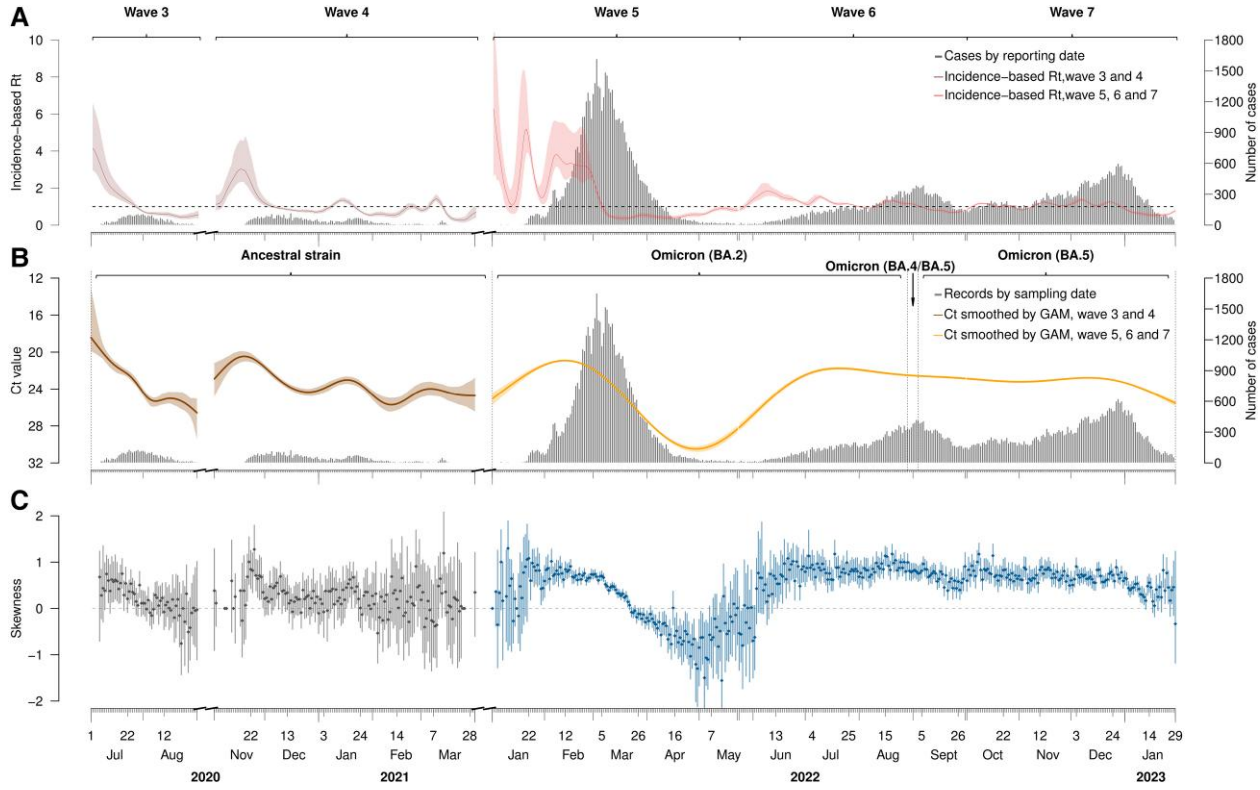

Figure 1. Temporal distribution of incidence-based R_t (estimated by case counts) and population-level Ct values (measured by daily mean and skewness). *A*, Locally confirmed COVID-19 cases with available Ct values by date of reporting and incidence-based R_t estimated by case counts. Black bars indicate daily case counts. Lines and shaded areas indicate the mean and 95% credible intervals 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 mean and 95% confidence intervals of Ct values estimated by a generalized additive model (GAM) during the third and fourth waves, which were dominated by the ancestral strain. Orange lines and shaded areas correspond to the fifth, sixth, and seventh waves, which were dominated by Omicron variants. *C*, Temporal distribution of Ct skewness. Dots and vertical lines represent the mean and 95% confidence intervals of daily Ct skewness. Ct, cycle threshold; R_t, effective reproductive number.

epidemic waves, except waves 4 and 6 (Table 2), with AUCs of 0.68 (95% CI, .60–.75; test set, wave 4), 0.98 (95% CI, .96–1.00; test set, wave 5), 0.62 (95% CI, .53–.70; test set, wave 6), and 0.80 (95% CI, .73–.88; test set, wave 7).

We validated our methods by training the model using 31-day peaks data from waves 4 to 7. The model demonstrated generally discriminative predictions, particularly in forecasting transmission in waves 3 and 5 (Table 2). The AUC of the predicted R_t for the Omicron variants–dominated wave 5 remained outstanding regardless of the training set used. Similarly, retrospective R_t estimates for the ancestral strain–dominated wave 3 indicated excellent, even outstanding, estimates with AUC values ranging from 0.80 to 0.92 for different training sets. However, the results occasionally showed suboptimal performance, especially when predicting R_t of plateaued waves 6 and 7. Furthermore, discriminative to nondiscriminative performances were observed when wave 5 was the training set, with AUCs of the predicted R_t being 0.53 (95% CI, .49–.57) for wave 6 and 0.49 (95% CI, .46–.53) for wave 7.

Impact of Severity on the Association Between Population-Level Ct Distribution and Incidence-Based R_t

There were more severe cases recorded during Omicron waves as compared with ancestral strain waves. Specifically, 40% of cases were retrospectively classified as severe in waves 5 and 7, as compared with <25% in waves 3 and 4 (Supplementary Table 3). As suggested, we retrospectively assessed Ct-based R_t estimates based on subsets of Ct values from cases with 2 distinct degrees of severity. Slightly distinct temporal distributions emerged in Ct values between mild to moderate and severe groups during waves 5 and 6 (Supplementary Figure 3a). For instance, Ct means diverged in February, April, and August 2022, coinciding with differing delays between symptom onset and sample collection for the 2 severity groups (Supplementary Figure 10c and 10d).

Results for the model trained in wave 3 suggested that using Ct values from mild to moderate or severe yielded only nondiscriminative AUCs (eg, 0.52 [95% CI, .48–.56] for wave 5, with severe cases only) as compared with using data from all cases (eg, 0.98 [95% CI, .96–1.00] for wave 5), and the model trained

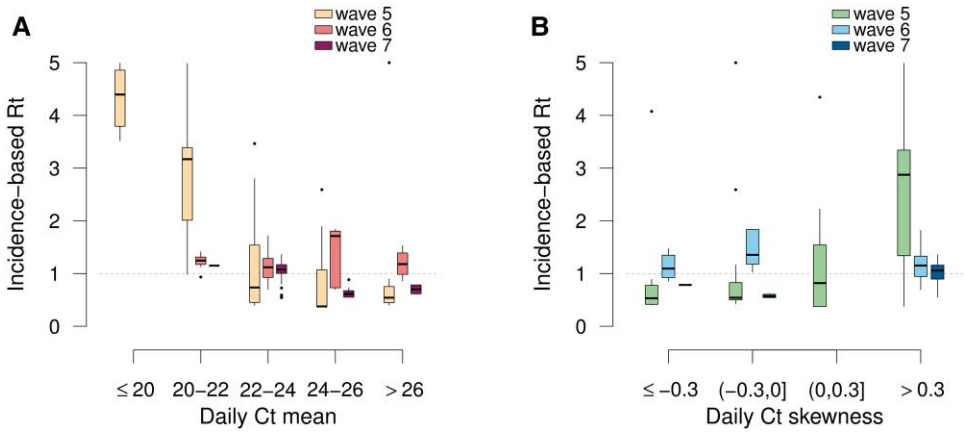

Figure 2. Distribution of incidence-based R_t under various intervals of Ct (panel *A*) mean or (panel *B*) skewness among Omicron waves. Plots indicate the median (line), IQR (box), 95% CI (error bars), and outliers (dots). Ct, cycle threshold; R_t, effective reproductive number.

in wave 4 show similar trend (Supplementary Table 5). In contrast, for the training periods with an increasing proportion of severe cases, using Ct from 1 severity group could yield higher AUCs as compared with using Ct from all cases. For instance, the AUC of the Ct-based R_t for wave 6 was 0.63 (95% CI, .55–.71; trained with mild to moderate cases from wave 5) as compared with 0.53 (95% CI, .49–.57; trained with all cases from wave 5). Finally, we retrospectively described the distribution of the interval between the first Ct sampling and the onset of severe cases, as depicted in Supplementary Figure 11, showing shorter intervals in the subsequent Omicron waves.

DISCUSSION

In this study, we demonstrated that the association between population viral load distribution and epidemics, as derived from the ancestral strain with minimal population immunity, remained highly informative during the Omicron subvariants–dominated waves in populations with significant preexisting immunity. In the previous study, reduced viral shedding durations were observed for vaccinated individuals and those infected with the Omicron variants; however, disparities in viral loads were minimal during the early stage of infection [7]. Therefore, our approach employing the initial Ct may be minimally affected by variants and vaccinations, consistent with previous studies using various measurements of transmission (eg, growth rate) [6, 27–30]. Our findings also suggested several circumstances in which the model exhibited suboptimal performance, such as amid sample representativeness or during plateaued local epidemics, thereby providing insights into broader model applicability.

For surveillance based primarily on symptoms and contact tracing [4], Ct samples could be delayed in case identification or sample collection, resulting in an inaccurate reflection of the overall viral load in the population. We observed lower population Ct values for severe cases in early wave 5 but higher values in late wave 5 and early wave 6 as compared with mild to moderate cases (Supplementary Figure 3), which coincided with observed disparities in delays from symptom onset to sample collection (Supplementary Figure 10). Our stratified analyses of sample severity profile also revealed a complex impact of severity on the Ct-based estimation of R_t, highlighting the importance of sample representativeness. During waves 3 and 4 in Hong Kong, when intensive surveillance and contact tracing were adopted amid low virus circulation in the community, models trained on all cases outperformed those trained only on mild to moderate cases, as they better represented the full spectrum of case exposure time distribution (Supplementary Figures 5–9). Conversely, when training the model with data from waves that have overall increases in the proportion of severe samples with delayed reporting, the model may inaccurately associate these changes in severity profile and reporting to changes in transmission, leading to reduced model performance (Supplementary Table 5). This observation is consistent with the shorter delays observed between initial Ct reports and clinical outcomes during Omicron waves as compared with the delays observed during waves 3 and 4 (Supplementary Figure 11). Note that our symptom-stratified analysis used retrospective data, where severe outcomes were ascertained months after the initial case confirmation. These outcomes would not be available in real time during the initial confirmation of cases.

Contact tracing was suspended, and RAT was implemented in February 2022 due to the huge impact on the medical system [9]. It was not until June 7 that cases with positive RAT results were required to undergo RT-qPCR again [12]. These policy changes were reflected in the differences in the percentages of RT-qPCR-confirmed cases and the percentages of RT-qPCR-positive cases with Ct values (Supplementary Table 2). Ct values were collected from RT-qPCR-confirmed cases admitted or quarantined in public hospital in Hong Kong [4,

Table 2. Model Performance Based on Different Training Periods to Estimate Ct-Based R_t in the Other 4 Waves

	Training Period									
	Wave 3[a] (19 Jul–18 Aug 2020)		Wave 4 (24 Nov–24 Dec 2020)		Wave 5 (21 Feb–23 Mar 2022)		Wave 6 (23 Aug–22 Sep 2022)		Wave 7 (19 Dec 2022–18 Jan 2023)	
	AUC	MAPE	AUC	MAPE	AUC	MAPE	AUC	MAPE	AUC	MAPE
Wave 3	0.94 (.86–1.00)	0.25	0.92 (.84–1.00)	0.28	0.80 (.70–.90)	3.45	0.91 (.83–1.00)	0.25	0.92 (.85–1.00)	0.27
Wave 4	0.68 (.60–.75)	0.29	0.69 (.61–.76)	0.30	0.67 (.59–.75)	16.40	0.70 (.62–.77)	0.32	0.71 (.64–.79)	0.37
Wave 5	0.98 (.96–1.00)	0.52	0.98 (.96–1.00)	0.47	0.96 (.92–1.00)	0.67	0.99 (.97–1.00)	0.56	0.98 (.96–1.00)	0.58
Wave 6	0.62 (.53–.70)	0.24	0.62 (.53–.70)	0.22	0.53 (.49–.57)	0.56	0.66 (.58–.74)	0.23	0.66 (.57–.75)	0.23
Wave 7	0.80 (.73–.88)	0.15	0.81 (.73–.88)	0.16	0.49 (.46–.53)	0.54	0.53 (.50–.56)	0.18	0.67 (.59–.75)	0.15
Overall[b]	0.78 (.74–.82)	0.29	0.83 (.79–.86)	0.27	0.53 (.49–.57)	6.00	0.69 (.65–.73)	0.33	0.80 (.76–.84)	0.37

Incidence-based *Rt* was natural log transformed.

Abbreviations: AUC, area under the receiver operating characteristic curve; Ct, cycle threshold; MAPE, mean absolute percentage error; R_t, effective reproductive number.

[a]Main model used to estimate Ct-based R_t.

[b]Overall: combined 4 test sets (excluded the wave where 31-day peak period was located) into 1 to calculate corresponding AUC and MAPE.

12]. During the Omicron waves, the surge in cases overwhelmed testing capacity and hospital resources, leading to a significant change in the availability of Ct values among confirmed cases. During this period, the incidence-based R_t was estimated by RT-qPCR–confirmed cases. Therefore, wave 5 was not recommended as a training set, which is why lower transmissibility was observed when the interval between symptom onset and sampling increased.

The accuracy of Ct-based R_t estimates for wave 6 was not optimal, irrespective of whether the model was trained with the data from the same wave or other waves. A possible reason could be the low and relatively consistent community transmission, as evidenced by the incidence-based R_t stably fluctuating around 1 and its Gini coefficient of 0.140 (95% CI, .125–.156; Supplementary Table 6). Simultaneously, fewer samples available to obtain Ct values led to increased uncertainties surrounding the R_t estimates [4]. In an effort to mitigate the effects of the limited sample size, we opted to fit the model using rolling Ct values instead of daily values on days with fewer than 30 or 60 records. Although the AUC and MAPE improved for wave 4 when trained on data from wave 3, suboptimal performance may arise in wave 6 because the moving average of Ct values could disrupt the alignment between the temporal distribution of Ct values and R_t (Supplementary Table 7). The results also indicated that the impact of the sample size on the Ct-based framework can be mitigated as long as the sample is representative [4]. Yet, increased AUCs for testing sets in waves 4 and 6 were observed when we excluded days with fewer than 30 or 60 records (Supplementary Table 8). Furthermore, the training period included for wave 6 coincided with the presence of multiple Omicron variants, including BA.2 and BA.4/BA.5. These variants were indicated to exhibit distinct viral kinetics and infectiousness [31, 32]; however, limited genotyping data pose challenges in evaluating the impact of the transition period when competing Omicron variants coexisted. Future study on the model incorporating strain-specific parameters, such as peak viral load and duration of RT-qPCR positivity, could minimize fluctuations related to the time since symptom onset. This approach would minimize the impact of the viral kinetics of circulating strains, providing a clearer reflection of the relationship between viral loads and epidemic transmission, thereby improving the numerical accuracy of the Ct-based framework.

Our work has several limitations. First, we were unable to further examine the effect of vaccines and variants due to the limited individual data. Second, the reduction in case ascertainment as nonpharmaceutical interventions were relaxed [9] and the potential bias toward reporting severe cases in the later epidemic waves may have affected the accuracy of R_t estimation based on case counts. Further research is needed to address this issue and refine the model accordingly. Additionally, it is worth noting that our model, requiring less computation efforts and crude metric AUC, enables robust binomial estimation of R_t values >1 or <1. Yet it may not afford absolute quantitative estimates, as demonstrated in the nowcast for wave 5 (Table 2, Supplementary Table 4, Supplementary Figure 1b).

Our study provides valuable insights into the potential of population-level Ct distribution as a predictive tool for timely assessment of transmission dynamics during waves characterized by variants dominating and population immunity shifting. These findings suggest the potential generalizability of this simplified framework across various settings and situations. However, it is important to exercise caution when applying our model to situations with limited Ct records, transitioning epidemic phases, or fluctuating sampling delay, as the model may perform suboptimally in these scenarios. Further research is required to investigate the reasons for the suboptimal scenarios and to assess the impact of different surveillance practices and viral characteristics on the association between population viral shedding and transmission, which would better inform the applicability of the Ct-based estimation framework.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We thank the Department of Health and Hospital Authority of the Food and Health Bureau of the Government of Hong Kong for providing the data for the analysis. We also thank Justin K. Cheung and Chloe S. Chui for their help in managing and collecting the data.

Author contributions. All authors meet the ICMJE criteria for authorship. B. Y. and B. J. C. conceived the study. Y. M., Y. L., W. X., E. H. Y. L., F. H., J. Y. W., P. W., and T. K. T. prepared the data. B. Y., Y. M., and Y. L. developed the model. Y. M. conducted the data analyses. Y. M., B. Y., T. K. T., and B. J. C. interpreted the results. Y. M. wrote the first draft of the article. All authors provided critical review and revision of the text and approved the final version.

Financial support. This work was supported by the Health and Medical Research Fund, Food and Health Bureau, Government of the Hong Kong Special Administrative Region (grant 22210552 to B. Y.); the Theme-Based Research Scheme (project T11-705/21-N to B. J. C.) and the General Research Fund (project 17100822 to T. K. T.) of the Research Grants Council of the Hong Kong SAR Government.

Potential conflicts of interest. B. J. C. consults for AstraZeneca, Fosun Pharma, GSK, Haleon, Moderna, Roche, Sanofi Pasteur, and Pfizer. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Tsang TK, Wu P, Lau EHY, Cowling BJ. Accounting for imported cases in estimating the time-varying reproductive number of coronavirus disease 2019 in Hong Kong. J Infect Dis **2021**; 224:783–7.
2. Ho F, Parag KV, Adam DC, et al. Accounting for the potential of overdispersion in estimation of the time-varying reproduction number. Epidemiology **2023**; 34:201–5.
3. Hay JA, Kennedy-Shaffer Lee, Kanjilal S, et al. Estimating epidemiologic dynamics from cross-sectional viral load distributions. Science **2021**; 373:eabh0635.
4. Lin Y, Yang B, Cobey S, et al. Incorporating temporal distribution of population-level viral load enables real-time estimation of COVID-19 transmission. Nat Commun **2022**; 13:1155.
5. Jones TC, Biele G, Mühlemann B, et al. Estimating infectiousness throughout SARS-CoV-2 infection course. Science **2021**; 373:eabi5273.
6. Sala E, Shah IS, Manissero D, et al. Systematic review on the correlation between SARS-CoV-2 real-time PCR cycle threshold values and epidemiological trends. Infect Dis Ther **2023**; 12:749–75.
7. Lin Y, Wu P, Tsang TK, 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 **2023**; 4:e722–31.
8. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol **2013**; 178: 1505–12.
9. Yang B, Lin Y, Xiong W, et al. Comparison of control and transmission of COVID-19 across epidemic waves in Hong Kong: an observational study. Lancet Reg Health West Pac **2024**; 43:100969.
10. Tsui ELH, Lui CSM, Woo PPS, 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 **2020**; 20:323.
11. Wong CKH, Lau KTK, Au ICH, 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 **2023**; 23:683–95.
12. Wong JY, Cheung JK, Lin 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**; 228:1231–9.
13. Becker NG, Watson LF, Carlin JB. A method of non-parametric back-projection and its application to AIDS data. Stat Med **1991**; 10:1527–42.
14. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med **2020**; 26:672–5.
15. Thompson RN, Stockwin JE, van Gaalen RD, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics **2019**; 29:100356.
16. Salje H, Cummings DAT, Rodriguez-Barraquer I, et al. Reconstruction of antibody dynamics and infection histories to evaluate dengue risk. Nature **2018**; 557:719–23.
17. Government of the Hong Kong Special Administrative Region. COVID-19 and flu express: local situation of COVID-19 activity (as of October 21, 2022). **2022**.

Available at: https://www.chp.gov.hk/files/pdf/local_situation_covid19_en_20221021.pdf. Accessed 27 March 2023.
18. Janssens A, Martens FK. Reflection on modern methods: revisiting the area under the ROC curve. Int J Epidemiol **2020**; 49:1397–403.
19. Dong J, Feng T, Thapa-Chhetry B, et al. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. Crit Care **2021**; 25:288.
20. Cowling BJ, Ali ST, Ng TWY, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health **2020**; 5:e279–88.
21. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol **2010**; 5:1315–6.
22. School of Public Health, The University of Hong Kong. Real-time dashboard. **2023**. Available at: https://covid19.sph.hku.hk. Accessed 27 March 2023.
23. Government of the Hong Kong Special Administrative Region. Latest situation of COVID-19 (as of August 31, 2022). **2022**. Available at: https://www.chp.gov.hk/files/pdf/local_situation_covid19_en.pdf. Accessed 28 July 2023.
24. Government of the Hong Kong Special Administrative Region. COVID-19 and flu express: local situation of COVID-19 activity (as of February 8, 2023). **2023**. Available at: https://www.chp.gov.hk/en/resources/29/100148.html. Accessed 28 July 2023.
25. McMenamin ME, Nealon J, Lin Y, 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 **2022**; 22:1435–43.
26. Government of the Hong Kong Special Administrative Region. Hong Kong vaccination dashboard. **2022**. Available at: https://www.covidvaccine.gov.hk/en/. Accessed 2 January 2022.
27. Tso CF, 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 **2021**; 7:e28265.
28. Stevens R, Pratama R, Naing Z, Condylios 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 **2022**; 54:800–2.
29. Mishra B, Ranjan J, Purushotham P, et al. High proportion of low cycle threshold value as an early indicator of COVID-19 surge. J Med Virol **2022**; 94:240–5.
30. Walker AS, Pritchard E, House T, et al. Ct threshold values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide variation across populations and over time. Elife **2021**; 10:e64683.
31. Wang Q, Guo Y, Iketani S, et al. Antibody evasion by SARS-CoV-2 omicron subvariants BA.2.12.1, BA.4 and BA.5. Nature **2022**; 608:603–8.
32. Hay JA, Kennedy-Shaffer L, Mina MJ. Viral loads observed under competing strain dynamics. medRxiv. Preprint posted online 30 July 2021. https://doi.org/10.1101/2021.07.27.21261224.