Preplanned Studies

Risk Factors Associated with the Spatiotemporal Spread of the SARS-CoV-2 Omicron BA.2 Variant — Shanghai Municipality, China, 2022

Wen Zheng[1]; Xiaowei Deng[1]; Cheng Peng[1]; Xuemei Yan[1]; Nan Zheng[1]; Zhiyuan Chen[1]; Juan Yang[1]; Marco Ajelli[2]; Juanjuan Zhang[1,#]; Hongjie Yu[1,3,#]

Summary

What is already known about this topic?

Previous studies have explored the spatial transmission patterns of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and have assessed the associated risk factors. However, none of these studies have quantitatively described the spatiotemporal transmission patterns and risk factors for Omicron BA.2 at the micro (within-city) scale.

What is added by this report?

This study highlights the heterogeneous spread of the 2022 Omicron BA.2 epidemic in Shanghai, and identifies associations between different metrics of spatial spread at the subdistrict level and demographic and socioeconomic characteristics of the population, human mobility patterns, and adopted interventions.

What are the implications for public health practice?

Disentangling different risk factors might contribute to a deeper understanding of the transmission dynamics and ecology of coronavirus disease 2019 and an effective design of monitoring and management strategies.

An Omicron BA.2 epidemic occurred in Shanghai, China in early March 2022. The objective of our study is to quantify the spatial spread of the epidemic across Shanghai subdistricts and identify risk factors. This study provides quantitative estimates of the epidemic arrival time, growth rate, and infection attack rate (IAR) as of May 31, 2022, and uses a generalized linear mixed effect model (GLMM) to explore their associations with demographic and socioeconomic characteristics of the population, human mobility, and interventions at the subdistrict level. We found that the epidemic growth rate was positively associated with the epidemic arrival time and subdistricts farther away from the (likely) origin of the outbreak had lower growth rates. The IAR was negatively correlated with the arrival time, distance from the initial outbreak location, subdistrict location, and booster coverage in the population aged 65 years and above; a positive association was found for population density and gross domestic product (GDP). This study highlights the role of the geographical structure of the city, human mobility, population characteristics, and adopted interventions in shaping the dynamics of the epidemic.

Shanghai is divided into 16 districts and 216 subdistricts. In the initial phase of the outbreak, grid management was implemented at the subdistrict level and entailed partial lockdown and mass nucleic acid screening for high-risk areas and non-high-risk areas. Afterward, eastern Shanghai entered a population-wide lockdown on March 28, and then the rest of Shanghai entered a lockdown phase on April 1 (Supplementary Figure S1, available in https://weekly.chinacdc.cn/). The city-wide lockdown was fully lifted on June 1, 2022.

Daily aggregated data on the number of infections and individual-level data of all severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections were extracted from multiple publicly available official data sources. The initial (identified) foci of the outbreak was the cultural activity center of Shiquan subdistrict in Putuo District, where a cluster of 14 SARS-CoV-2 positive individuals was detected starting from March 1, 2022 (*1*).

To describe the time course of the Omicron outbreak in Shanghai, we estimated the following three indicators at the subdistrict level: 1) epidemic arrival time (i.e., the date of the first confirmed infection in a subdistrict), 2) IAR (i.e., the cumulative number of reported infections in a subdistrict divided by the total population in that subdistrict), and 3) epidemic growth rate.

To explore potential risk factors associated with the epidemic arrival time, growth rate, and IAR across

subdistricts, we included several covariates that belong to four general categories: demographic characteristics, socioeconomic characteristics, human mobility, and interventions (Supplementary Table S1, available in https://weekly.chinacdc.cn/). The arrival time of the epidemic represents a response variable when measuring the spread of the infection; however, we also considered it as an explanatory variable when exploring its association with the epidemic growth rate and IAR.

A correlation analysis was conducted to assess collinearities between the independent variables. We built a GLMM model to estimate the proportion of variance in the response variables ascribable to intra- and inter-district variation. The significance level was set to 0.1 for candidate variable selection, and 0.05 for multivariate regression. To test whether the random model was appropriately chosen, we also estimated spatial autocorrelation between residuals using Moran's *I* statistic. To quantify the uncertainty of model selection, a generalized estimating equation (GEE) model accounting for spatial clustering was used in a sensitivity analysis. The detailed statistical methods are presented in the Supplementary Material. All the analyses were performed in R 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria).

As of May 31, 2022, a total of 626,840 SARS-CoV-2 infections had been reported in 99.54% of the Shanghai subdistricts. High heterogeneity in the spatial distribution of infections was found across subdistricts, with 27.78% of the subdistricts accounting for more than 70% of all infections (Supplementary Figure S2, available in https://weekly.chinacdc.cn/).

The spatial spread of the epidemic showed a clear spatial trend from the city center to adjacent areas, and a continuous spread toward suburban and rural areas. The spatial distribution of the arrival time was highly heterogeneous, with 35.19%, 41.67%, and 97.69% of the subdistricts reporting infections within the first week, second week, and a month, respectively.

We analyzed the correlation between the epidemic arrival time and the geographical distance from the initial outbreak location. The regression model showed that compared to the geographic and effective distances, the pre-epidemic flow of travelers showed a slightly weaker correlation with the epidemic arrival time (Figure 1).

By fitting a linear regression model to the logarithm of the daily number of new confirmed infections from February 26 to April 1, 2022, the overall epidemic growth rate for Shanghai was estimated to be 0.23 per day [95% confidence interval (*CI*): 0.22–0.25]. Excluding 6 subdistricts reporting no infections before the lockdown and 49 subdistricts with R^2<0.6, as well as 3 subdistricts with only two data points, we analyzed the estimated growth rates for the remaining 158 subdistricts. The growth rate was lognormal-distributed, with a range of 0.06 to 0.39, which was positively associated with the arrival time of the epidemic (Figure 2A–2C).

The results of the univariate analysis were reported in the Supplementary Table S2 (available in https://weekly.chinacdc.cn/). The final selected model showed that the arrival time positively correlated with the growth rate of the epidemic [odds ratio (*OR*): 1.03, 95% *CI*: 1.02–1.04]. Subdistricts located in the suburban ring (*OR*: 0.85, 95% *CI*: 0.73–0.98) and outside the suburban ring (*OR*: 0.59, 95% *CI*: 0.49–0.70) were associated with a significantly lower epidemic growth rate (Figure 2D). The residuals did not show significant spatial autocorrelation with Moran's *I* analysis. The results were robust after removing the outliers. We obtained similar results with

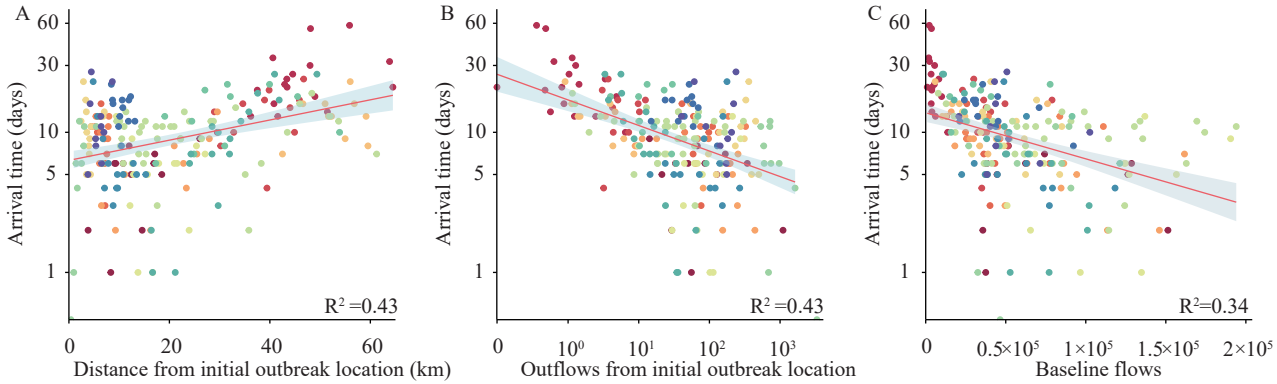


FIGURE 1. Scatter plots of the correlation between epidemic arrival time and (A) geographical distance, (B) effective distance, and (C) baseline flows.
Note: Dots in the scatter plot were colored by different districts.

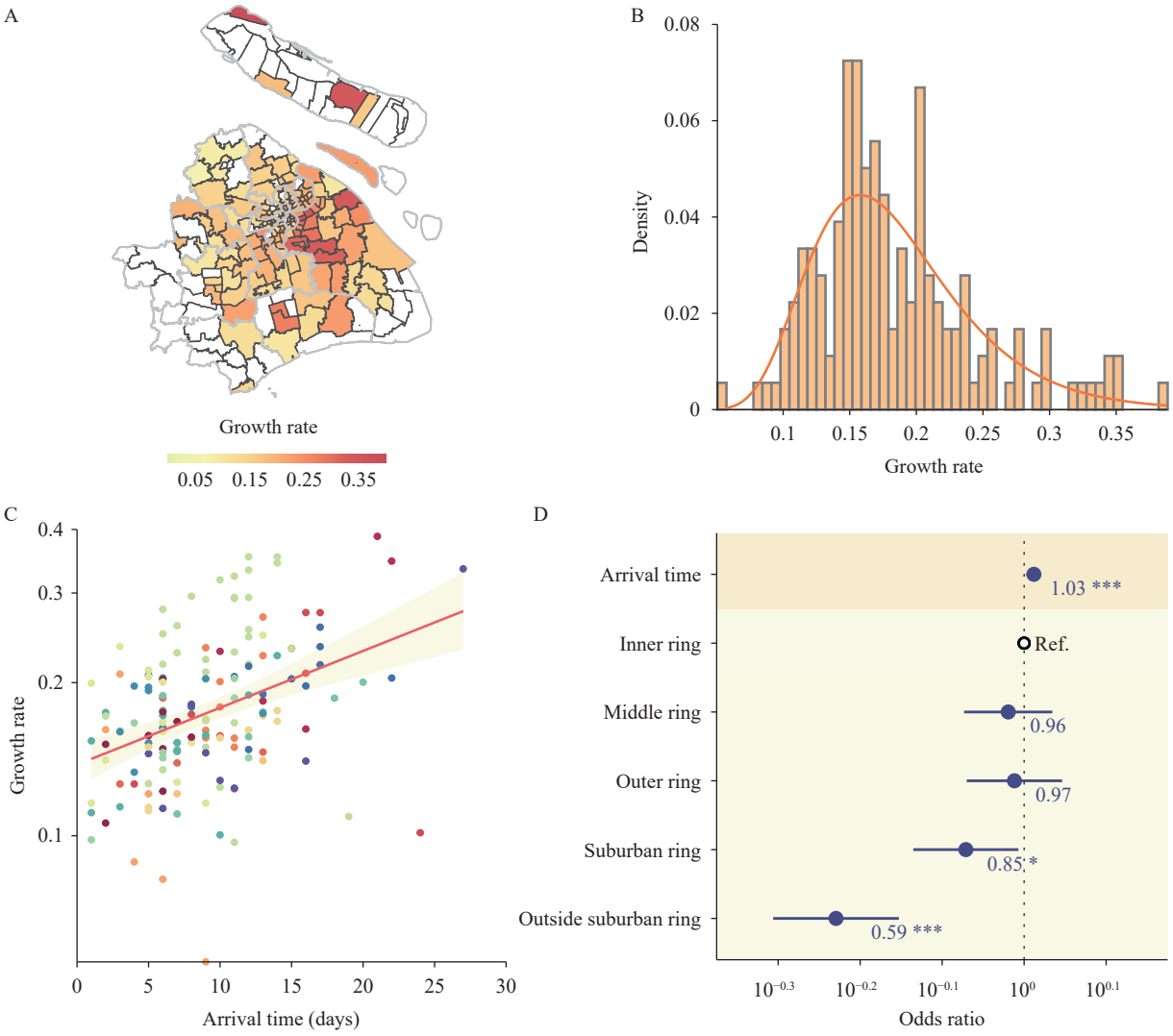


FIGURE 2. Epidemic growth rate and its associated factors. (A) Geographical distribution of growth rates for 158 subdistricts. (B) Distribution of the estimated epidemic growth rate (per day) by subdistrict and fit of a log-normal distribution. (C) Scatter plot of the epidemic growth rate and arrival time. (D) Factors associated with the growth rate.
Note: In panel A, of the 216 subdistricts, 6 subdistricts reporting no infections before the lockdown, 49 subdistricts with an R^2<0.6, and 3 subdistricts with only two data points for estimating their growth rates, were excluded from the regression; the excluded subdistricts are shown in white. In panel C, dots were colored by different districts. In panel D, dots and lines indicate point estimates and 95% confidence intervals of the odds ratio. Odds ratio was calculated as exponentiated regression coefficients. Numbers on the side of the dots indicate the numerical value of the point estimate.
Abbreviation: Ref.=reference category.
* indicates *P*-value<0.05;
** indicates *P*-value<0.01;
*** indicates *P*-value<0.001.

a GEE model (Supplementary Table S3, available in https://weekly.chinacdc.cn/).

As of May 31, 2022, the overall IAR in Shanghai was estimated to be 2.42%. However, the IAR was highly heterogeneous across subdistricts, ranging from 0 to 13.75%. The epidemic arrival time was significantly associated with the IAR. The final selected model showed that the arrival time was negatively correlated with the IAR (*OR*: 0.59, 95% *CI*: 0.46–0.75). Among the investigated demographic characteristics, population density was positively associated with the IAR (*OR*: 1.38, 95% *CI*: 1.20–1.60). Among the socioeconomic characteristics, subdistricts that were farther apart from the initial outbreak location were associated with a significantly lower IAR (*OR*: 0.96, 95% *CI*: 0.95–0.98). GDP at the district level positively correlated with the IAR (*OR*: 1.51, 95% *CI*: 1.23–1.85). Compared to

subdistricts located in the inner ring, subdistricts located farther away from the inner ring were significantly associated with lower IARs (*OR* for middle ring: 0.51, 95% *CI*: 0.35–0.73; *OR* for outer ring: 0.43, 95% *CI*: 0.32–0.58; *OR* for suburban ring: 0.16, 95% *CI*: 0.11–0.24; *OR* for outside suburban ring: 0.05, 95% *CI*: 0.03–0.08). Among the vaccine-related covariates, booster coverage for people aged 65 years and above was associated with a significantly lower IAR (*OR*: 0.73, 95% *CI*: 0.55–0.96, Figure 3). Moran's *I* for the residuals showed no significant spatial autocorrelation. The results were robust after removing the outliers, and similar results were obtained with a GEE model (Supplementary Table S4, available in https://weekly.chinacdc.cn/).

DISCUSSION

This study highlights the heterogeneous spread of the 2022 Omicron BA.2 epidemic in Shanghai, and identifies associations between different metrics of spatial spread at the subdistrict level and demographic and socioeconomic characteristics of the population, human mobility patterns, and adopted interventions.

The identified (likely) foci of the outbreak was the cultural activity center of the Shiquan subdistrict in Putuo District. However, we could not rule out the possibility that the Omicron outbreak might have

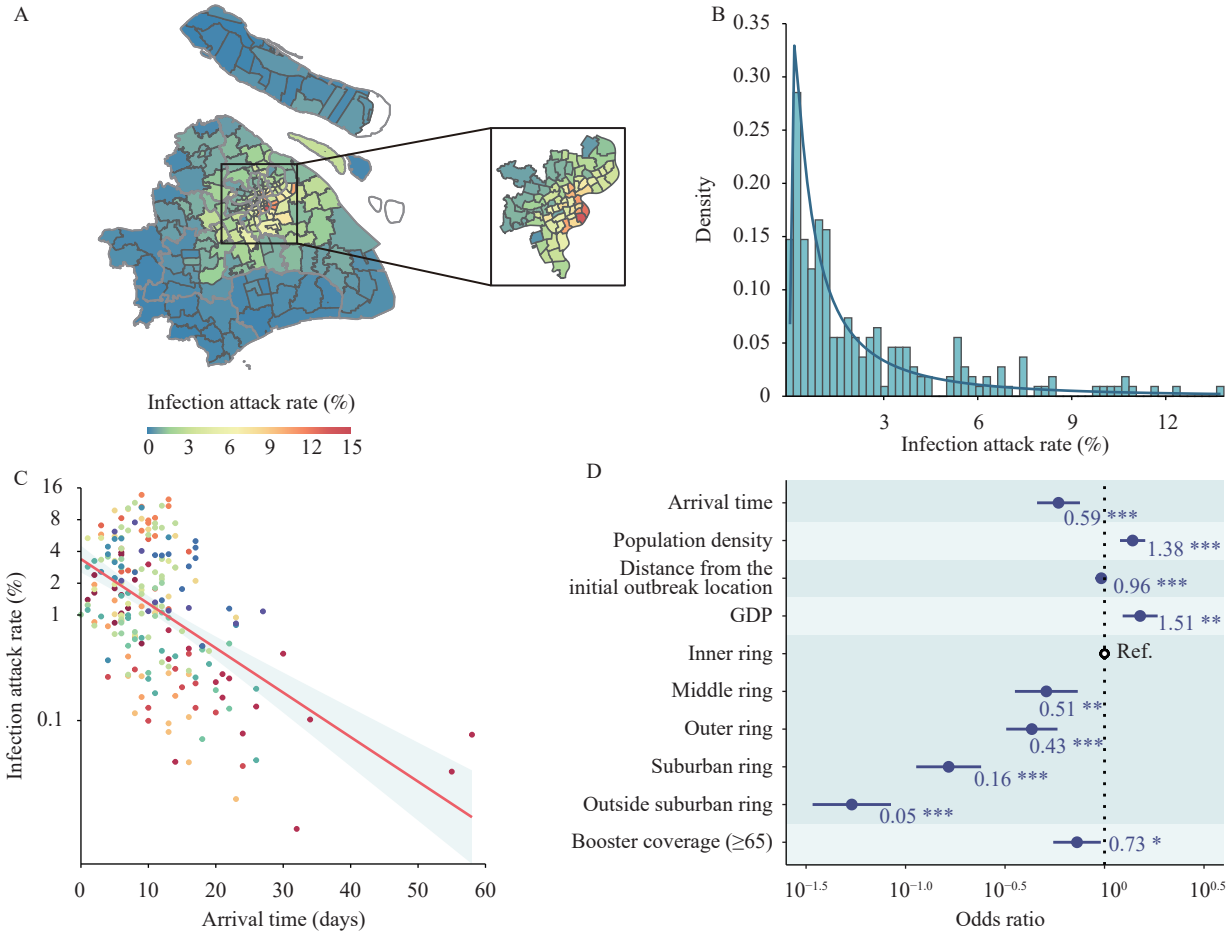


FIGURE 3. Infection attack rate and its associated factors. (A) Geographical distribution of the infection attack rates at the subdistrict level as of May 31, 2022. (B) Distribution of the infection attack rate by subdistrict and fit of a log-normal distribution. (C) Scatter plot of the infection attack rate and arrival time. (D) Factors associated with the infection attack rate.
Note: In panel C, dots were colored by different districts. In panel D, dots and lines indicate point estimates and 95% confidence intervals of the odds ratio. Odds ratio was calculated as exponentiated regression coefficients. Numbers on the side of the dots indicate the numerical value of the point estimates.
Abbreviation: GDP=gross domestic product; Ref.=reference category.
* indicates *P*-value<0.05;
** indicates *P*-value<0.01;
*** indicates *P*-value<0.001.

originated from multiple sources that resulted in simultaneous transmission chains prior to the identification of the first local transmission event.

Our findings suggest that the subdistricts with stronger connections to the initial outbreak location had higher chances of being reached by the epidemic early on, which is consistent with observations for the 2009 H1N1 influenza pandemic and 2003 SARS epidemic (*2*). The epidemic growth rate was positively associated with the arrival time of the epidemic, suggesting that the targeted interventions implemented in high-risk areas were insufficient to slow down transmission (*3*).

Human mobility is generally considered to be key in determining the risk of infection and the spread of epidemics (*4*–*5*). However, in our multivariate regression models, we found that pre-epidemic population flows were not significantly associated with the IAR or growth rate. This could be explained by its strong collinearity with the epidemic arrival time, which ultimately had a strong impact on determining the type and timing of adoption of control measures (*6*).

Subdistricts with a higher GDP were found to have a higher IAR, consistent with a previous study (*7*). This indicates that subdistricts with a higher GDP trend to have more factories and enterprises, causing more gathering and higher risk of transmission accordingly.

Initially, the outbreak spread in and around the inner ring. Previous literature also found that the COVID-19 pandemic in the United States was characterized by a geographically localized mosaic of transmission along an urban-rural gradient (*8*–*9*), suggesting that geographic distance may play an important role in SARS-CoV-2 spread. Finally, the public health impact of COVID-19 vaccines has already been widely discussed in the literature (*10*–*13*) and our study confirms previous evidence.

Our study suffers from limitations that are rooted in the uncertainty and fragmentary nature of publicly available sources, such as a high level of missing data for key variables, such as the date of symptom onset. Additionally, the population flows were provided by China Unicom and thus may suffer from the limitation intrinsic of mobile phone data. Moreover, we cannot exclude the possibility that there are other potential risk factors that were not considered in our study (e.g., housing conditions and meteorological factors). Meteorological factors may play an important role to explain the heterogeneity in the temporal and spatial spread of infectious diseases, but we did not include them here as the variation may be very limited across the small study location and the short study period. Finally, this study does not provide causal relationships, but only provides associations between different metrics of the epidemic spread with a set of indicators.

In conclusion, this study provides a quantitative description of the spatiotemporal spread of the Omicron BA.2 variant in Shanghai at the subdistrict level. Our findings highlight the role of the geographical structure of the city, human mobility, socioeconomic characteristics of the population, and adopted interventions in shaping the dynamics of the epidemic. Disentangling these factors might contribute to a deeper understanding the transmission dynamics and ecology of COVID-19 and guide the design of monitoring and management strategies.

Funding: Supported by the Key Program of the National Natural Science Foundation of China (82130093 to H.Y.) and Shanghai Rising-Star Program (22QA1402300 to J.Z.).

doi: 10.46234/ccdcw2023.018

[#] Corresponding authors: Juanjuan Zhang, zhangjuan@fudan.edu.cn; Hongjie Yu, yhj@fudan.edu.cn.

[1] School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai Municipality, China; [2] Laboratory for Computational Epidemiology and Public Health, Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN, USA; [3] Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai Municipality, China.

Submitted: October 27, 2022; Accepted: January 28, 2023

REFERENCES

1. Chen ZY, Deng XW, Fang LQ, Sun KY, Wu YP, Che TL, et al. Epidemiological characteristics and transmission dynamics of the outbreak caused by the SARS-CoV-2 Omicron variant in Shanghai, China: a descriptive study. Lancet Reg Health West Pac 2022;29:100592. http://dx.doi.org/10.1016/j.lanwpc.2022.100592.
2. Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. Science 2013;342(6164):1337 – 42. http://dx.doi.org/10.1126/science.1245200.
3. Hay JA, Kissler SM, Fauver JR, Mack C, Tai CG, Samant RM, et al. Quantifying the impact of immune history and variant on SARS-CoV-2 viral kinetics and infection rebound: a retrospective cohort study. Elife 2022;11:e81849. http://dx.doi.org/10.7554/eLife.81849.
4. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Soldan VP, Kochel TJ, Kitron U, et al. The role of human movement in the transmission of vector-borne pathogens. PLoS Negl Trop Dis 2009;3(7):e481. http://dx.doi.org/10.1371/journal.pntd.0000481.
5. Doorley R, Berke A, Noyman A, Alonso L, Ribó J, Arroyo V, et al. Mobility and COVID-19 in andorra: country-scale analysis of high-resolution mobility patterns and infection spread. IEEE J Biomed Health Inform 2022;26(1):183 – 93. http://dx.doi.org/10.1109/JBHI.

2021.3121165.
6. Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020;368(6490):493 – 7. http://dx.doi.org/10.1126/science.abb4218.
7. Liu SR, Qin YC, Xie ZX, Zhang JF. The spatio-temporal characteristics and influencing factors of COVID-19 spread in Shenzhen, China-an analysis based on 417 cases. Int J Environ Res Public Health 2020;17(20):7450. http://dx.doi.org/10.3390/ijerph17207450.
8. Susswein Z, Valdano E, Brett T, Rohani P, Colizza V, Bansal S. Ignoring spatial heterogeneity in drivers of SARS-CoV-2 transmission in the US will impede sustained elimination. medRxiv 2021. http://dx.doi.org/10.1101/2021.08.09.21261807.
9. Souch JM, Cossman JS, Hayward MD. Interstates of infection: preliminary investigations of human mobility patterns in the COVID-19 pandemic. J Rural Health 2021;37(2):266 – 71. http://dx.doi.org/10.1111/jrh.12558.
10. Tan ST, Park HJ, Rodríguez-Barraquer I, Rutherford GW, Bibbins-Domingo K, Schechter R, et al. COVID-19 vaccination and estimated public health impact in California. JAMA Netw Open 2022;5(4):e228526. http://dx.doi.org/10.1001/jamanetworkopen.2022.8526.
11. Watson OJ, Barnsley G, Toor J, Hogan AB, Winskill P, Ghani AC. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. Lancet Infect Dis 2022;22(9):1293 – 302. http://dx.doi.org/10.1016/S1473-3099(22)00320-6.
12. Suthar AB, Wang J, Seffren V, Wiegand RE, Griffing S, Zell E. Public health impact of COVID-19 vaccines in the US: observational study. BMJ 2022;377:e069317. http://dx.doi.org/10.1136/bmj-2021-069317.
13. Bubar KM, Middleton CE, Bjorkman KK, Parker R, Larremore DB. SARS-CoV-2 transmission and impacts of unvaccinated-only screening in populations of mixed vaccination status. Nat Commun 2022;13(1):2777. http://dx.doi.org/10.1038/s41467-022-30144-7.

SUPPLEMENTARY MATERIAL

Generalized Linear Mixed Effect Model

Arrival time. We built a generalized linear mixed effect model (GLMM) to estimate the proportion of variance in the response variables attributable to intra- and inter-district variation. For the arrival time, we used a GLMM with random intercept effect and log link function to assess its association with geographical distance from the initial outbreak location. We used the concept of "effective distance" (*1*), wherein the distance between locations depends on the strength of their link; in our case, the strength of the link is measured as the outflows from initial outbreak location.

Epidemic growth rate. Similarly, we used a GLMM with random intercept, random slope, and log link function to explore whether and to what extent the arrival time correlates with the epidemic growth rate, while controlling for other covariates. To select which explanatory variables to include in the final model, we first used univariate regression for candidate variable selection, then we run multivariate regressions. Finally, we performed a forward stepwise model selection based on Akaike's Information Criterion (AIC) and likelihood ratio test.

The specification of the final GLMM for the epidemic growth rate is the following:

$$g(\mu_{ij}) = \alpha + \beta_1 arrival\_time_{ij} + \beta_{2j} ring_{ij} + u_j$$

where g is a log link function; i represents the subdistrict; j represents the district; α represents the intercept; $arrival\_time_{ij}$ and $ring_{ij}$ denote the fixed effects of the arrival time and ring where the subdistrict is located; u_j represents the district-specific random intercept effects; β_{2j} represents the district-specific random slope effects; and $\mu_{ij} = E(Y_{ij} | u_j, \beta_{2j})$ is the mean of the response variable (i.e., the epidemic growth rate) Y_{ij} for a given value of the random effects.

Infection attack rate. To explore driving factors associated with the infection attack rate, we used the same GLMM. After model selection, the specification of the final GLMM for IAR is the following:

$$g(\mu_{ij}) = \alpha + \beta_{1j} arrival\_time_{ij} + \beta_2 density_{ij} + \beta_{3j} ring_{ij} + \beta_4 distance_{ij} + \beta_5 GDP_{ij} + \beta_6 booster\_65_{ij} + u_j$$

where g is a log link function; i represents the subdistrict; j represents the district; α represents the intercept; $arrival\_time_{ij}$, $density_{ij}$, $ring_{ij}$, $distance_{ij}$, GDP_{ij}, and booster_65_{ij} denote the fixed effects of the arrival time, population density, ring where the subdistrict is located, distance from the initial outbreak location, GDP, and booster vaccination coverage of people aged 65 years and above; u_j represents the district-specific random intercept effects; β_{1j} and β_{3j} represents the district-specific random slope effects; and $\mu_{ij} = E(Y_{ij} | u_j, \beta_{1j}, \beta_{3j})$ is the mean of the response variable (i.e., the infection attack rate) Y_{ij} for a given value of the random effects.

Observations with a Cook's Distance greater than 20 times the mean value were considered outliers and excluded from the analysis. Odd ratios were calculated by exponentiating the coefficients and confidence interval from the regression results. Diagnostics were performed to assess regression assumptions.

Generalized Estimating Equation Model

In addition to the GLMM, we also built a generalized estimating equation (GEE) model to regress the epidemic growth rate and infection attack rate. The GEE model relies on a similar specification of the initial GLM fitting, but with no random effects. To further account for spatial clustering and possible correlation structure, GEE uses u_j to define the clustering structure of the data with a working correlation matrix that defines the correlation within each cluster (i.e., district). We provide the results for the exchangeable correlation matrix, but independent and unstructured matrices were explored as well and gave very similar results.

SUPPLEMENTARY TABLE S1. Definition and data sources for potential risk factors.

Type	Factors	Level	Data source
Demographic characteristics*			
	1. Population density (×1,000 people/km^2)	Subdistrict	6th and 7th Census (*2*–*3*)
	2. Proportion of people aged 65 years old and over (%)	Subdistrict	6th and 7th Census (*2*–*3*)
	3. Ratio of population with household registration to the effective population (%)	Subdistrict	Shanghai Statistics Year Book (*4*)
Socioeconomic characteristics			
	4. Distance from the initial outbreak location, i.e., cultural activity center in Shiquan subdistrict of Putuo District (km)	Subdistrict	Amap (*5*)
	5. Gross domestic product (GDP) (×100 million CNY)	District	Shanghai Statistics Year Book (*4*)
	6. Coverage of green area (%)	District	Shanghai Statistics Year Book (*4*)
	7. Ring where the subdistrict is located ● Inner ring: subdistricts with more than half of the area within the Inner-city Elevated Beltway; ● Middle ring: subdistricts between the Inner-city Elevated Beltway and the Middle Ring Road; ● Outer ring: subdistricts between the Middle Ring Road and the Outer Ring Road, i.e., S20; ● Suburban ring: subdistricts between the S20 Road and the Highway around Shanghai City, i.e., G1501; ● Outside the suburban ring: subdistricts outside the G1501 Road	Subdistrict	Public sources
Human behavior†			
	8. Baseline flows during pre-epidemic (×1,000 trips): daily average population flows (inflows, outflows and inner flows) for a given subdistrict in the last week of February, i.e., between February 21 and February 27, 2022.	Subdistrict	China Unicom (*6*)
	9. Outflows from the initial outbreak location (i.e., Shiquan subdistrict) to other 215 subdistricts during pre-epidemic (×1,000 trips): daily average outflows between February 21 and February 27, 2022	Subdistrict	China Unicom (*6*)
Vaccine and non-pharmaceutical interventions			
	10. Vaccine coverage (%)§ ● Primary vaccination coverage of total population ● Booster vaccination coverage of total population ● Primary vaccination coverage of people aged 65 years old and over ● Booster vaccination coverage of people aged 65 years old and over	Subdistrict	Public sources and internal report
	11. Whether a given subdistrict was classified as high-risk area or not, during grid management phase between March 16 and March 27, 2022	Subdistrict	Public sources
	12. Lockdown time of eastern and western Shanghai, defined as subdistricts east and west of the Huangpu River ● Eastern Shanghai: March 28 ● Western Shanghai: April 1	Subdistrict	Public sources
	13. Reduction in daily population flows after lockdown (%): the subtraction of daily average flows during early lockdown (between April 1 and April 7, 2022) from the baseline flows and the division by the baseline	Subdistrict	China Unicom[a] (*6*)

Abbreviation: CNY=Chinese Yuan.

* Subdistrict-level population data after 2017 were derived from the 7th National Census of China and the latest reports by local authorities. For the subdistricts with unavailable population data after 2017, the subdistrict-level population data for 2020 were inferred from the population size of each district in 2020 and the population proportion of each subdistrict in the Sixth National Census in 2010.

† The population flow data is provided by one of the largest national mobile carriers in China, China Unicom, and is aggregated based on all users' mobile phone activity records across the city, including geographic location. We then aggregated the daily inflows, outflows, and internal flows at the subdistrict level.

§ The numerator is vaccinated individuals, and the denominator is census population. If floating population who was vaccinated were counted in the numerator, it may result in coverage exceeding 100%. Besides, census population for some subdistricts is not up to date, possibly leading to overestimation of coverage. Thus, the coverage would be truncated to 100%, if exceeding 100%.

SUPPLEMENTARY TABLE S2. Univariate regression for growth rate and infection attack rate.

Effect	Epidemic growth rate			Infection attack rate (%)		
	Estimate	Pr(>|z|)	95% *CI*	Estimate	Pr(>|z|)	95% *CI*
Population density	1.01	<0.001***	(1.003, 1.01)	2.16	<0.001***	(1.95, 2.39)
Proportion of people aged ≥65	1.07	0.271	(0.95, 1.20)	0.99	0.140	(0.97, 1.004)
Ratio of population with household registration to the effective population	1.00	0.918	(0.93, 1.06)	2.05	<0.001***	(1.67, 2.52)
Distance from initial outbreak location	0.91	0.036*	(0.83, 0.99)	0.93	<0.001***	(0.92, 0.94)
GDP	1.03	0.591	(0.92, 1.17)	2.23	0.050	(1.08, 4.60)
Coverage of green area	1.00	0.763	(0.98, 1.03)	0.93	0.401	(0.81, 1.08)
Ring where the subdistrict is located						
Inner ring	Ref	–	–	Ref	–	–
Middle ring	0.94	0.373	(0.82, 1.08)	0.46	0.009**	(0.30, 0.72)
Outer ring	0.94	0.424	(0.81, 1.09)	0.40	0.005**	(0.26, 0.63)
Suburban ring	0.82	0.021*	(0.70, 0.97)	0.15	<0.001***	(0.09, 0.24)
Outside suburban ring	0.61	<0.001***	(0.50, 0.74)	0.05	<0.001***	(0.03, 0.08)
Baseline flows	1.02	0.593	(0.95, 1.10)	1.86	<0.001***	(1.53, 2.27)
Outflows from initial outbreak location	1.04	0.089	(0.99, 1.08)	1.54	<0.001***	(1.39, 1.72)
Primary coverage of total population	0.99	<0.001***	(0.99, 0.996)	0.44	0.074	(0.18, 1.08)
Booster coverage of total population	0.99	<0.001***	(0.98, 0.996)	0.11	<0.001***	(0.06, 0.22)
Primary coverage of people aged ≥65	1.00	0.381	(1.00, 1.00)	0.81	0.370	(0.52, 1.26)
Booster coverage of people aged ≥65	1.00	0.726	(0.99, 1.00)	0.74	0.150	(0.50, 1.11)
High-risk area	0.93	0.141	(0.85, 1.02)	1.70	<0.001***	(1.29, 2.25)
Lockdown time						
March 28	–	–	–	Ref	–	–
April 1	–	–	–	4.53	<0.001***	(2.35, 8.76)
Reduction in flows after lockdown	–	–	–	1.05	<0.001***	(1.04, 1.07)
Arrival time	1.02	<0.001***	(1.01, 1.03)	0.53	0.004**	(0.37, 0.75)

Abbreviation: GDP=gross domestic product; *CI*=confidence intervals; Ref=reference.
* *P*<0.05;
** *P*<0.01;
*** *P*<0.001.

SUPPLEMENTARY TABLE S3. Generalized estimating equation model for growth rate.

Effect	Exchangeable			Independence			Unstructured		
	Estimate	Pr(>|z|)	95% *CI*	Estimate	Pr(>|z|)	95% *CI*	Estimate	Pr(>|z|)	95% *CI*
Intercept	0.13	<0.001***	(0.12, 0.15)	0.13	<0.001***	(0.12, 0.15)	0.13	<0.001***	(0.12, 0.15)
Arrival time	1.03	<0.001***	(1.02, 1.04)	1.03	<0.001***	(1.02, 1.04)	1.03	<0.001***	(1.02, 1.04)
Ring where the subdistrict is located									
Inner ring	Ref	–	–	Ref	–	–	Ref	–	–
Middle ring	1.00	0.986	(0.88, 1.14)	1.00	0.983	(0.88, 1.14)	1.00	0.985	(0.88, 1.14)
Outer ring	1.05	0.253	(0.94, 1.26)	1.09	0.254	(0.94, 1.26)	1.09	0.253	(0.94, 1.26)
Suburban ring	1.01	0.865	(0.90, 1.14)	1.01	0.869	(0.90, 1.14)	1.01	0.866	(0.90, 1.14)
Outside suburban ring	0.79	0.061	(0.61, 1.01)	0.79	0.060	(0.61, 1.01)	0.79	0.061	(0.61, 1.01)

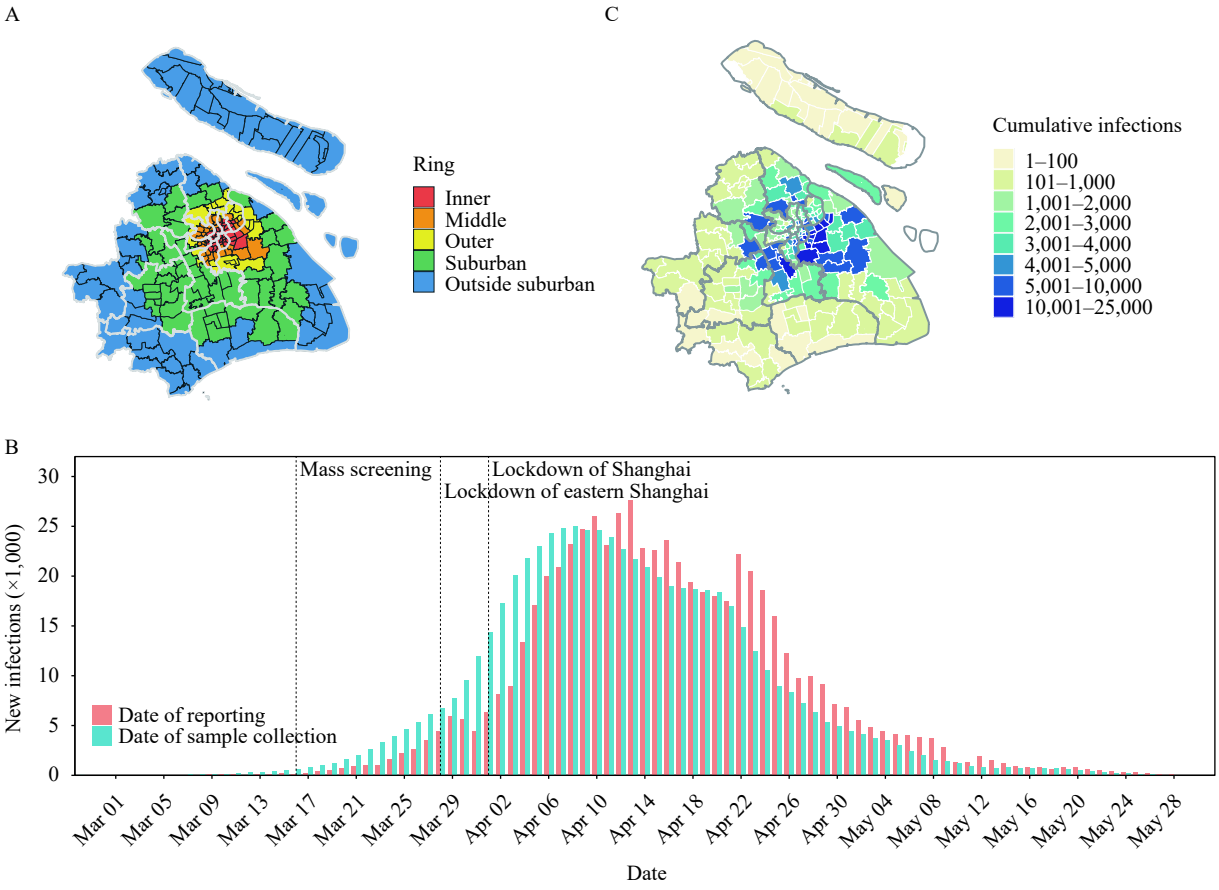
Note: Total *N*=158. df=152.
Abbreviation: *CI*=confidence intervals; Ref=reference.
* *P*<0.05;
** *P*<0.01;
*** *P*<0.001.

SUPPLEMENTARY TABLE S4. Generalized estimating equation model for infection attack rate.

Effect	Exchangeable			Independence			Unstructured		
	Estimate	Pr(>|z|)	95% *CI*	Estimate	Pr(>|z|)	95% *CI*	Estimate	Pr(>|z|)	95% *CI*
Intercept	0.01	<0.001***	(0, 0.03)	0.01	<0.001***	(0.01, 0.03)	0.01	<0.001***	(0, 0.04)
Arrival time	0.75	0.013*	(0.59, 0.94)	0.74	0.009**	(0.59, 0.93)	0.73	0.008**	(0.58, 0.92)
Population density	1.51	0.001**	(1.20, 1.91)	1.57	<0.001***	(1.26, 1.96)	1.65	<0.001***	(1.29, 2.10)
Distance from the initial outbreak location	1.01	0.653	(0.98, 1.03)	1.01	0.536	(0.98, 1.03)	1.01	0.442	(0.98, 1.04)
GDP	1.35	<0.001***	(1.17, 1.56)	1.35	<0.001***	(1.20, 1.53)	1.33	0.001**	(1.13, 1.58)
Ring where the subdistrict is located									
Inner ring	Ref	–	–	Ref	–	–	Ref	–	–
Middle ring	0.45	<0.001***	(0.34, 0.59)	0.42	<0.001***	(0.32, 0.56)	0.44	<0.001***	(0.31, 0.63)
Outer ring	0.46	<0.001***	(0.36, 0.58)	0.43	<0.001***	(0.34, 0.55)	0.44	<0.001***	(0.34, 0.58)
Suburban ring	0.18	<0.001***	(0.13, 0.24)	0.17	<0.001***	(0.13, 0.23)	0.17	<0.001***	(0.12, 0.24)
Outside suburban ring	0.06	<0.001***	(0.04, 0.09)	0.05	<0.001***	(0.04, 0.08)	0.06	<0.001***	(0.04, 0.09)
Booster coverage of people aged ≥ 65 years old	0.67	0.001**	(0.53, 0.84)	0.67	0.002**	(0.53, 0.86)	0.68	0.002**	(0.54, 0.87)

Note: Total *N*=216. d*f*=198.
Abbreviation: *CI*=confidence intervals; Ref=reference.
* *P*<0.05;
** *P*<0.01;
*** *P*<0.001.

SUPPLEMENTARY FIGURE S1. Geographic division of eastern and western Shanghai.
Note: Eastern and western Shanghai are naturally separated by the Huangpu River (blue layer). Specifically, eastern Shanghai contained the districts of Pudong New Area, Fengxian, Jinshan, Chongming, as well as partial Minhang and Songjiang; while the rest areas were grouped into western Shanghai.



SUPPLEMENTARY FIGURE S2. Temporal dynamics and geographical distribution of confirmed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections. (A) Visualization of the ring where the subdistrict is located. (B) Geographical distribution of confirmed SARS-CoV-2 infections at the subdistrict level. (C) Daily number of new confirmed infections by date of reporting and by date of sample collection.
Note: Shanghai is divided into 16 districts (light grey boundary) and 216 subdistricts (black boundary) shown in panel A. The colored area corresponds to the ring where the subdistrict is located.

REFERENCES

1. Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. Science 2013;342(6164):1337 – 42. http://dx.doi.org/10.1126/science.1245200.
2. Office of the Leading Group of the State Council for the Seventh National Population Cencus. China population census yearbook 2020. Beijing: China Statistics Press. 2022. https://book.kongfz.com/13826/5510107594/. (In Chinese).
3. Population Census Office Under the State Council Department of Population, Employment Statistics National Bureau of Statistics. Tabulation on the 2010 population census of the people's republic of China. Beijing: China Statistics Press. 2012. https://book.kongfz.com/351691/4922881913/. (In Chinese).
4. Shanghai Municipal Bureau of Statistics, Survey office of the National Bureau of Statistics in Shanghai. Shanghai statistical yearbook 2021. Beijing: China Statistics Press. 2021. https://book.kongfz.com/8654/4665739962/. (In Chinese).
5. Amap. https://ditu.amap.com/. [2022-6-29]. (In Chinese).
6. SmartSteps. SmartSteps. http://www.smartsteps.com/. [2022-6-29]. (In Chinese).