

Original Paper

Nonrepresentativeness of Human Mobility Data and its Impact on Modeling Dynamics of the COVID-19 Pandemic: Systematic Evaluation

Chuchu Liu^{1,2}, PhD; Petter Holme^{3,4}, Prof Dr; Sune Lehmann⁵, Prof Dr; Wenchuan Yang², PhD; Xin Lu², Prof Dr

¹School of Economics and Management, Changsha University of Science and Technology, Changsha, China

²College of Systems Engineering, National University of Defense Technology, Changsha, China

³Department of Computer Science, Aalto University, Espoo, Finland

⁴Center for Computational Social Science, Kobe University, Kobe, Japan

⁵Department of Applied Mathematics and Computer Science, Technical University of Denmark, Copenhagen, Denmark

Corresponding Author:

Xin Lu, Prof Dr

College of Systems Engineering

National University of Defense Technology

No 137 Yanwachi Street

Changsha, 410073

China

Phone: 86 18627561577

Email: xin.lu.lab@outlook.com

Abstract

Background: In recent years, a range of novel smartphone-derived data streams about human mobility have become available on a near-real-time basis. These data have been used, for example, to perform traffic forecasting and epidemic modeling. During the COVID-19 pandemic in particular, human travel behavior has been considered a key component of epidemiological modeling to provide more reliable estimates about the volumes of the pandemic's importation and transmission routes, or to identify hot spots. However, nearly universally in the literature, the representativeness of these data, how they relate to the underlying real-world human mobility, has been overlooked. This disconnect between data and reality is especially relevant in the case of socially disadvantaged minorities.

Objective: The objective of this study is to illustrate the nonrepresentativeness of data on human mobility and the impact of this nonrepresentativeness on modeling dynamics of the epidemic. This study systematically evaluates how real-world travel flows differ from census-based estimations, especially in the case of socially disadvantaged minorities, such as older adults and women, and further measures biases introduced by this difference in epidemiological studies.

Methods: To understand the demographic composition of population movements, a nationwide mobility data set from 318 million mobile phone users in China from January 1 to February 29, 2020, was curated. Specifically, we quantified the disparity in the population composition between actual migrations and resident composition according to census data, and shows how this nonrepresentativeness impacts epidemiological modeling by constructing an age-structured SEIR (Susceptible-Exposed-Infected-Recovered) model of COVID-19 transmission.

Results: We found a significant difference in the demographic composition between those who travel and the overall population. In the population flows, 59% (n=20,067,526) of travelers are young and 36% (n=12,210,565) of them are middle-aged ($P<.001$), which is completely different from the overall adult population composition of China (where 36% of individuals are young and 40% of them are middle-aged). This difference would introduce a striking bias in epidemiological studies: the estimation of maximum daily infections differs nearly 3 times, and the peak time has a large gap of 46 days.

Conclusions: The difference between actual migrations and resident composition strongly impacts outcomes of epidemiological forecasts, which typically assume that flows represent underlying demographics. Our findings imply that it is necessary to measure and quantify the inherent biases related to nonrepresentativeness for accurate epidemiological surveillance and forecasting.

(JMIR Form Res 2024;8:e55013) doi: [10.2196/55013](https://doi.org/10.2196/55013)

KEYWORDS

human mobility; data representativeness; population composition; COVID-19; epidemiological modeling

Introduction

With large-scale empirical data (eg, mobile phone records, GPS data, and location-based social network data) becoming available with increasingly fine spatial and temporal resolution [1], quantitative studies on individual and collective mobility patterns have flourished in the past few years [2-6]. These developments have offered advances with respect to understanding migratory flows, traffic forecasting, urban planning, and epidemic modeling [7-10]. The ongoing COVID-19 pandemic has further intensified discussions on how to optimally use human mobility research to support outbreak responses and nonpharmaceutical interventions (eg, contact tracing) [11-14].

The representativeness of data sets used to infer real-world human mobility, however, has typically not been explicitly incorporated in such analyses. This is potentially troubling as representativeness is known to be especially poor for socially disadvantaged minorities such as low-income groups, women, children, and older people. For example, it has been confirmed that individuals' probability of travel is not randomly or equally distributed, and there is significant heterogeneity when comparing the travel patterns of different demographic groups [15-17]. For example, women are more localized than men in their movements and visit fewer locations in regions such as Latin America, Bangladesh, and sub-Saharan Africa [18,19]. In the specific case of epidemic outbreaks, low-income individuals are not necessarily able to limit their exposure to a circulating virus by reducing mobility and must continue, for example, commuting behavior to remain employed. Thus, this group is subject to a substantially higher probability of becoming infected in an epidemic than higher-income groups [20]. Further, different contact rates across age groups have been observed in COVID-19 incidence cases [21], and higher COVID-19 infection rates among disadvantaged racial and socioeconomic groups have been observed in multiple studies [22,23]. It has also been argued that including information about demographic heterogeneity in human mobility patterns, for example, by combining demographically stratified travel data with epidemiology research, would make epidemiological models more robust [24].

While it is widely recognized that rich new data sources can provide near-real-time information about human mobility [25] and powerful input into models that estimate imported cases using regional mobility information when modeling pathogen transmission, the state-of-the-art models do not consider data representativeness. This is typically because for privacy concerns, most data sets are not disaggregated demographically. Instead, relevant information on demographic features and social relationships is traditionally collected by censuses and other surveys [26-28].

As we argue below, however, simply considering the population demographics at the origin of a trip does not represent the traveling population.

To systematically evaluate how real-world travel flows differ from census-based estimations, we use an aggregated and anonymized data set collected from 318 million mobile phones. Specifically, we quantified the disparity in the population composition between actual migrations and resident composition according to census data and found significant differences. We then investigated how this nonrepresentativeness impacts epidemiological modeling. The aim of this study is to illustrate the nonrepresentativeness of data on human mobility and the impact of this nonrepresentativeness on modeling dynamics of the COVID-19 pandemic.

Methods

Data Description

In China, a total of 847 million Chinese people use mobile phones to surf the internet, accounting for 99.1% of the total netizens. The penetration rate of mobile phone usage among the population aged 15-65 years is almost 100%, providing extensive coverage and high representativeness for the national population. To understand the demographic composition of population movements, we collected nationwide mobility data from 318 million mobile phone users in China from January 1 to February 29, 2020. All population flow data were aggregated on the basis of users' geographic locations and demographic characteristics (eg, gender and age). To enhance extrapolation and representation of the population, a machine learning method was used to extrapolate the data to all users of the entire network, which also agrees well with the official population statistics ($R^2=0.98$; [Multimedia Appendix 1](#) [2,29-31]).

Epidemiological Modeling

To illustrate the impact of data nonrepresentativeness on modeling dynamics of the COVID-19 epidemic, we constructed an age-structured SEIR (Susceptible-Exposed-Infected-Recovered) model of COVID-19 transmission developed by Prem et al [32]. In fitting this age-mixing transmission model with heterogeneous contact rates between age groups [33], the differential age composition of traveling people and the overall national population were input as alternative parameters. By comparing model outputs, we measured the bias caused by the data nonrepresentativeness of demographic composition in forecasting epidemic dynamics.

Ethical Considerations

The study data provided by the operator were anonymized (without personally identifying information) and aggregated at the city level. As no individual study was carried out, no ethical approval was required to undertake this scoping study.

Results

Demographic Heterogeneity Among Traveling Individuals

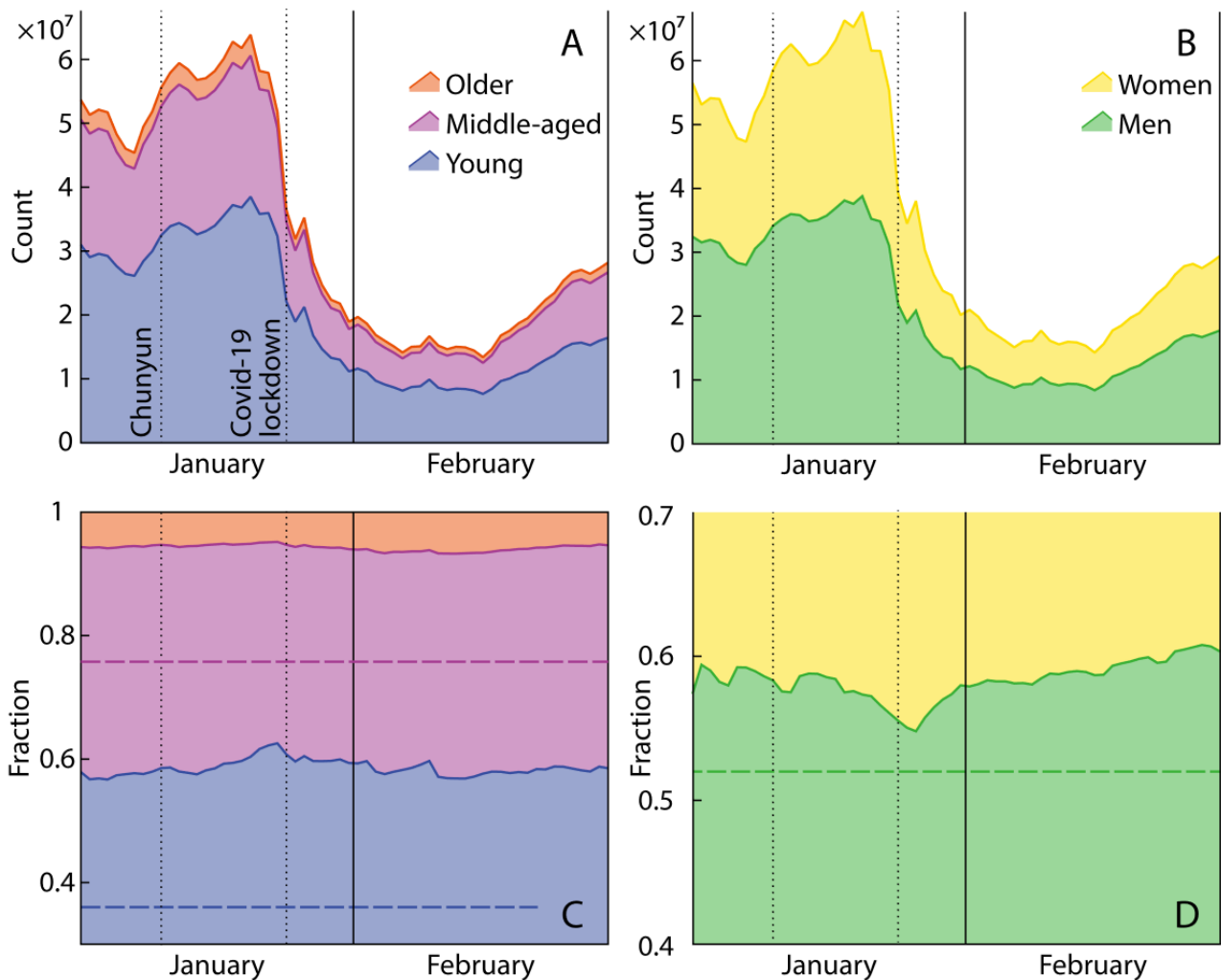
For our analysis, we draw on a unique data set from China. China is an ideal location to study representativeness in mobile

phone data because of its very high smartphone penetration. In China, the penetration rate of mobile phone usage among the population aged 15-65 years is almost 100%, providing extensive coverage and high representativeness for the national population [29]. We estimate the full national mobility at the city level by extrapolating from 318 million mobile phone users (see [Multimedia Appendix 1 \[2,29-31\]](#) for details).

Our comparison reveals a marked difference between the overall population composition and those who travel. Hereinafter, we define “young” individuals as those in their 20s-30s, “middle-aged” individuals as those in their 40s-50s, and “older adults” as those older than 60 years. Specifically, we found that the majority of population flows within China are generated by men and young people. Although mobility behavior fluctuated strongly across our observation period ([Figures 1A and 1B](#)) and is affected by temporal factors such as weekdays and holidays,

the composition of travelers did not change significantly over different periods ([Figures 1C and 1D](#)). In the population flows, 59% ($n=20,067,526$) of travelers are young and 36% ($n=12,210,565$) of them are middle-aged (as children generally do not have mobile phones, all ratios are calculated with minors younger than 18 years having been excluded). This ratio is completely different from what we observed in the overall adult population composition of China (about 1084 million in total) [30], where 36% ($n=390$ million) of individuals are young and 40% ($n=430$ million) of them are middle-aged. Furthermore, daily male travelers constitute approximately 59% ($n=20,858,026$) of the total number of traveling individuals, which is greater than the overall proportion of men (51.2%, $n=721$ million). Compared to men, women travel less often, but we found that when women travel, they tend to move slightly further than men, with 175 km traveled per person in an average intercity trip, compared to 170 km for men ($P<.001$).

Figure 1. Profiles of the intercity movements extracted from mobile phone data between January 1 and February 29, 2020, in China. (A) and (B) show the daily number of travelers for different age and gender groups; (C) and (D) show the respective ratios. Dashed horizontal lines denote the composition of respective groups in the latest 7th census. As children generally do not have mobile phones, the proportions of young (20-39 years), middle-aged (40-59 years), and older people (≥ 60 years) add up to 100%.



Bias From Data Nonrepresentativeness

Since individual mobility is the primary reason for the spatial diffusion of an epidemic, it is important to directly explore how

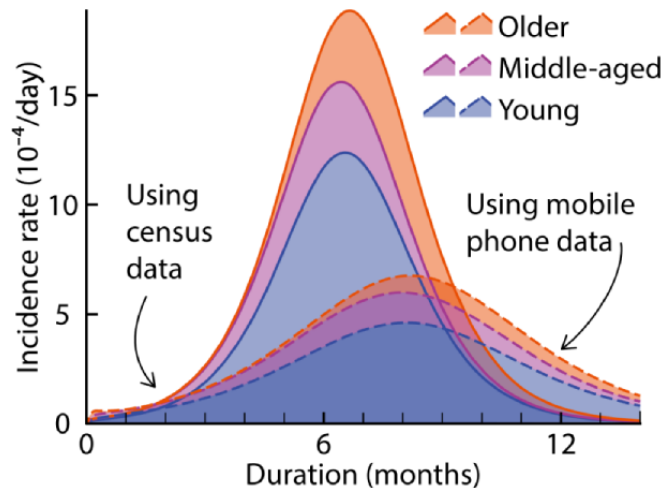
the demographic heterogeneity of human migration behaviors impacts our ability to forecast the spatial behavior of epidemics.

By fitting an age-structured transmission model [32,33], and including differential age composition as an input parameter, we measured the possible bias caused by the

nonrepresentativeness of data on the traveling population in modeling epidemic dynamics. Feeding the composition of travelers and composition from the census data separately into the model, we found that the predicted number of infected individuals had a striking bias: the maximum number of daily infections in these 2 populations differ by nearly 3 times (521 infected individuals among a total of 1 million people for the composition of travelers and 1521 infected individuals for the census population), and their peak time had a large gap of 46 days. Although older adults are the most susceptible population,

the 2 infection rates among older people collected from mobility data and census data deviate strongly (Figure 2). Further, while the predicted cumulative number of confirmed cases do gradually stabilize late in the epidemic, the gap between the results of the 2 models is nonnegligible (with a deviation of around 79.5%) with respect to infection volumes. Failing to include information about age and gender structure in real-world human mobility is thus likely to introduce considerable biases in epidemiological studies, especially in the early phase of an epidemic outbreak caused by imported cases [24].

Figure 2. Dynamics of the incidence rates among different groups predicted by the age-structured model. Solid lines indicate the incidence rates of different age groups from census data, and dashed lines indicate the incidence rates by using traveling data from mobile phones.



Discussion

By comparing mobility traces from mobile phone users to census data, our study has highlighted a number of striking differences in the demographic composition of those who travel with respect to the overall population. For example, we found that 59% ($n=20,067,526$) of travelers are young and 36% ($n=12,210,565$) of them are middle-aged, which is completely different from the composition of the adult population in China, where 36% of people are young and 40% of them are middle-aged. The travel probability and travel distance between men and women were also significantly different. This realization is especially important in the case of epidemic forecasting, and increased awareness of this issue in the scientific community has the potential to improve not only epidemiological models but also our overall understanding of possible biases when inferring human mobility from cell phone data and the representational issues of mobility data. It is important to emphasize that while China is an ideal place to study representativeness, our findings about which fraction of individuals compose the population of travelers are specific to China. The fraction of young, middle-aged, old, male, and female individuals who travel is likely to depend on a range of factors and can be expected to be different in different countries.

Nonetheless, the realization that understanding the representativeness of mobility data is crucial for epidemic monitoring and forecasting is generalizable. Thus, our results imply that when generalizing results from population mobility analysis, these differences should be included in the analysis to avoid potential biases caused by data nonrepresentativeness. For example, in the case of the COVID-19 pandemic, as travelers often have a higher probability of infection, the transmission risk among men and youth could be a promising focus for COVID-19 prevention.

In the recent Omicron waves, imported infections represented the majority of cases in China, and most COVID-19–positive individuals had a travel history to high-risk areas such as Shanghai [34]. As presymptomatic and asymptomatic pathogen carriers can travel to a foreign country and initiate the spread of COVID-19 even when there is no community transmission, human migration behaviors are promising candidates to incorporate into epidemiological models. Our findings emphasize that focusing on the representativeness of mobility data is essential for more sophisticated modeling approaches to capture key mechanisms of epidemic propagation. In future work, we intend to further explore how to accurately quantify the inherent biases related to data nonrepresentativeness for accurate epidemiological surveillance and forecasting.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grants 72025405, 72088101, 72001211, and 72301285), the National Social Science Foundation of China (grants 22ZDA102), the Hunan Science and Technology Plan Project (grants 2020TP1013), the Natural Science Foundation of Hunan Province (grants 2024JJ6069 and 2023JJ40685), and the Innovation

Team Project of Colleges in Guangdong Province (grants 2020KCXTD040). P.H. was supported by JSPS KAKENHI (grants JP 21H04595).

Data Availability

Deidentified data and code used in the analysis are available upon reasonable request from the corresponding author.

Authors' Contributions

XL designed the study. CL and XL analyzed the data. CL, PH, SL, XL, and WY contributed to the interpretation of the results and drafted the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Underrepresented data of the population flow.

[\[DOCX File , 984 KB-Multimedia Appendix 1\]](#)

References

1. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, et al. Human mobility: Models and applications. *Physics Reports*. Mar 2018;734:1-74. [doi: [10.1016/j.physrep.2018.01.001](https://doi.org/10.1016/j.physrep.2018.01.001)]
2. Tan S, Lai S, Fang F, Cao Z, Sai B, Song B, et al. Mobility in China, 2020: a tale of four phases. *Natl Sci Rev*. Nov 2021;8(11):nwab148. [FREE Full text] [doi: [10.1093/nsr/nwab148](https://doi.org/10.1093/nsr/nwab148)] [Medline: [34876997](https://pubmed.ncbi.nlm.nih.gov/34876997/)]
3. Hou X, Gao S, Li Q, Kang Y, Chen N, Chen K, et al. Intracounty modeling of COVID-19 infection with human mobility: assessing spatial heterogeneity with business traffic, age, and race. *Proc Natl Acad Sci U S A*. Jun 15, 2021;118(24):e2020524118. [FREE Full text] [doi: [10.1073/pnas.2020524118](https://doi.org/10.1073/pnas.2020524118)] [Medline: [34049993](https://pubmed.ncbi.nlm.nih.gov/34049993/)]
4. Schlosser F, Maier BF, Jack O, Hinrichs D, Zachariae A, Brockmann D. COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proc Natl Acad Sci U S A*. Dec 29, 2020;117(52):32883-32890. [FREE Full text] [doi: [10.1073/pnas.2012326117](https://doi.org/10.1073/pnas.2012326117)] [Medline: [33273120](https://pubmed.ncbi.nlm.nih.gov/33273120/)]
5. Lu X, Tan J, Cao Z, Xiong Y, Qin S, Wang T, et al. Mobile phone-based population flow data for the COVID-19 outbreak in Mainland China. *Health Data Sci*. Jun 18, 2021;2021:9796431. [FREE Full text] [doi: [10.34133/2021/9796431](https://doi.org/10.34133/2021/9796431)] [Medline: [36405355](https://pubmed.ncbi.nlm.nih.gov/36405355/)]
6. Xiong C, Hu S, Yang M, Luo W, Zhang L. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. *Proc Natl Acad Sci U S A*. Nov 03, 2020;117(44):27087-27089. [FREE Full text] [doi: [10.1073/pnas.2010836117](https://doi.org/10.1073/pnas.2010836117)] [Medline: [33060300](https://pubmed.ncbi.nlm.nih.gov/33060300/)]
7. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*. Jun 29, 2020;582(7812):389-394. [doi: [10.1038/s41586-020-2284-y](https://doi.org/10.1038/s41586-020-2284-y)] [Medline: [32349120](https://pubmed.ncbi.nlm.nih.gov/32349120/)]
8. Kraemer MUG, Yang C, Gutierrez B, Wu C, Klein B, Pigott DM, Open COVID-19 Data Working Group, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*. May 01, 2020;368(6490):493-497. [FREE Full text] [doi: [10.1126/science.abb4218](https://doi.org/10.1126/science.abb4218)] [Medline: [32213647](https://pubmed.ncbi.nlm.nih.gov/32213647/)]
9. Chen P, Liu R, Aihara K, Chen L. Autoreservoir computing for multistep ahead prediction based on the spatiotemporal information transformation. *Nat Commun*. Sep 11, 2020;11(1):4568. [FREE Full text] [doi: [10.1038/s41467-020-18381-0](https://doi.org/10.1038/s41467-020-18381-0)] [Medline: [32917894](https://pubmed.ncbi.nlm.nih.gov/32917894/)]
10. Liu R, Zhong J, Hong R, Chen E, Aihara K, Chen P, et al. Predicting local COVID-19 outbreaks and infectious disease epidemics based on landscape network entropy. *Sci Bull (Beijing)*. Nov 30, 2021;66(22):2265-2270. [doi: [10.1016/j.scib.2021.03.022](https://doi.org/10.1016/j.scib.2021.03.022)] [Medline: [36654453](https://pubmed.ncbi.nlm.nih.gov/36654453/)]
11. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci Adv*. Jun 05, 2020;6(23):eabc0764. [FREE Full text] [doi: [10.1126/sciadv.abc0764](https://doi.org/10.1126/sciadv.abc0764)] [Medline: [32548274](https://pubmed.ncbi.nlm.nih.gov/32548274/)]
12. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*. Sep 04, 2020;585(7825):410-413. [FREE Full text] [doi: [10.1038/s41586-020-2293-x](https://doi.org/10.1038/s41586-020-2293-x)] [Medline: [32365354](https://pubmed.ncbi.nlm.nih.gov/32365354/)]
13. Xia J, Yin K, Yue Y, Li Q, Wang X, Hu D, et al. Impact of human mobility on COVID-19 transmission according to mobility distance, location, and demographic factors in the Greater Bay Area of China: population-based study. *JMIR Public Health Surveill*. Apr 26, 2023;9:e39588. [FREE Full text] [doi: [10.2196/39588](https://doi.org/10.2196/39588)] [Medline: [36848228](https://pubmed.ncbi.nlm.nih.gov/36848228/)]
14. Li Z, Li X, Porter D, Zhang J, Jiang Y, Olatosi B, et al. Monitoring the spatial spread of COVID-19 and effectiveness of control measures through human movement data: proposal for a predictive model using big data analytics. *JMIR Res Protoc*. Dec 18, 2020;9(12):e24432. [FREE Full text] [doi: [10.2196/24432](https://doi.org/10.2196/24432)] [Medline: [33301418](https://pubmed.ncbi.nlm.nih.gov/33301418/)]

15. Adhikari S, Pantaleo NP, Feldman JM, Ogedegbe O, Thorpe L, Troxel AB. Assessment of community-level disparities in coronavirus disease 2019 (COVID-19) infections and deaths in large US metropolitan areas. *JAMA Netw Open*. Jul 01, 2020;3(7):e2016938. [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.16938](https://doi.org/10.1001/jamanetworkopen.2020.16938)] [Medline: [32721027](https://pubmed.ncbi.nlm.nih.gov/32721027/)]
16. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. *Nat Hum Behav*. May 30, 2020;4(5):460-471. [doi: [10.1038/s41562-020-0884-z](https://doi.org/10.1038/s41562-020-0884-z)] [Medline: [32355299](https://pubmed.ncbi.nlm.nih.gov/32355299/)]
17. Weill JA, Stigler M, Deschenes O, Springborn MR. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proc Natl Acad Sci U S A*. Aug 18, 2020;117(33):19658-19660. [FREE Full text] [doi: [10.1073/pnas.2009412117](https://doi.org/10.1073/pnas.2009412117)] [Medline: [32727905](https://pubmed.ncbi.nlm.nih.gov/32727905/)]
18. Grantz KH, Meredith HR, Cummings DAT, Metcalf CJE, Grenfell BT, Giles JR, et al. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat Commun*. Sep 30, 2020;11(1):4961. [FREE Full text] [doi: [10.1038/s41467-020-18190-5](https://doi.org/10.1038/s41467-020-18190-5)] [Medline: [32999287](https://pubmed.ncbi.nlm.nih.gov/32999287/)]
19. Sinha I, Sayeed AA, Uddin D, Wesolowski A, Zaman SI, Faiz MA, et al. Mapping the travel patterns of people with malaria in Bangladesh. *BMC Med*. Mar 04, 2020;18(1):45. [FREE Full text] [doi: [10.1186/s12916-020-1512-5](https://doi.org/10.1186/s12916-020-1512-5)] [Medline: [32127002](https://pubmed.ncbi.nlm.nih.gov/32127002/)]
20. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*. Jan 10, 2021;589(7840):82-87. [doi: [10.1038/s41586-020-2923-3](https://doi.org/10.1038/s41586-020-2923-3)] [Medline: [33171481](https://pubmed.ncbi.nlm.nih.gov/33171481/)]
21. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID COVID-19 working group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med*. Aug 16, 2020;26(8):1205-1211. [doi: [10.1038/s41591-020-0962-9](https://doi.org/10.1038/s41591-020-0962-9)] [Medline: [32546824](https://pubmed.ncbi.nlm.nih.gov/32546824/)]
22. Pareek M, Bangash MN, Pareek N, Pan D, Sze S, Minhas JS, et al. Ethnicity and COVID-19: an urgent public health research priority. *Lancet*. May 2020;395(10234):1421-1422. [doi: [10.1016/s0140-6736\(20\)30922-3](https://doi.org/10.1016/s0140-6736(20)30922-3)]
23. Chowkwanyun M, Reed AL. Racial health disparities and Covid-19 — caution and context. *N Engl J Med*. Jul 16, 2020;383(3):201-203. [doi: [10.1056/nejmp2012910](https://doi.org/10.1056/nejmp2012910)]
24. Buckee C, Noor A, Sattenspiel L. Thinking clearly about social aspects of infectious disease transmission. *Nature*. Jul 30, 2021;595(7866):205-213. [doi: [10.1038/s41586-021-03694-x](https://doi.org/10.1038/s41586-021-03694-x)] [Medline: [34194045](https://pubmed.ncbi.nlm.nih.gov/34194045/)]
25. Buckee CO, Balsari S, Chan J, Crosas M, Dominici F, Gasser U, et al. Aggregated mobility data could help fight COVID-19. *Science*. Apr 10, 2020;368(6487):145-146. [doi: [10.1126/science.abb8021](https://doi.org/10.1126/science.abb8021)] [Medline: [32205458](https://pubmed.ncbi.nlm.nih.gov/32205458/)]
26. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*. Mar 25, 2008;5(3):e74. [FREE Full text] [doi: [10.1371/journal.pmed.0050074](https://doi.org/10.1371/journal.pmed.0050074)] [Medline: [18366252](https://pubmed.ncbi.nlm.nih.gov/18366252/)]
27. Karaca-Mandic P, Georgiou A, Sen S. Assessment of COVID-19 hospitalizations by race/ethnicity in 12 states. *JAMA Intern Med*. Jan 01, 2021;181(1):131-134. [FREE Full text] [doi: [10.1001/jamainternmed.2020.3857](https://doi.org/10.1001/jamainternmed.2020.3857)] [Medline: [32804192](https://pubmed.ncbi.nlm.nih.gov/32804192/)]
28. Rubin D, Huang J, Fisher BT, Gasparrini A, Tam V, Song L, et al. Association of social distancing, population density, and temperature with the instantaneous reproduction number of SARS-CoV-2 in counties across the United States. *JAMA Netw Open*. Jul 01, 2020;3(7):e2016099. [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.16099](https://doi.org/10.1001/jamanetworkopen.2020.16099)] [Medline: [32701162](https://pubmed.ncbi.nlm.nih.gov/32701162/)]
29. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health*. May 2020;5(5):e261-e270. [doi: [10.1016/s2468-2667\(20\)30073-6](https://doi.org/10.1016/s2468-2667(20)30073-6)]
30. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*. Jun 26, 2020;368(6498):1481-1486. [FREE Full text] [doi: [10.1126/science.abb8001](https://doi.org/10.1126/science.abb8001)] [Medline: [32350060](https://pubmed.ncbi.nlm.nih.gov/32350060/)]
31. Baidu. URL: <https://qianxi.baidu.com/#/> [accessed 2024-05-24]
32. Number of mobile cell phone subscriptions in China from December 2020 to December 2023. Statista. URL: <https://www.statista.com/statistics/278204/china-mobile-users-by-month/>, [accessed 2022-06-02]
33. The seventh national census data from national bureau of statistics in China. National Bureau of Statistics. URL: http://www.stats.gov.cn/tjsj/tjgb/rkpcgb/qgrkpcgb/202106/t20210628_1818823.html, [accessed 2022-05-12]
34. Zhang J, Tan S, Peng C, Xu X, Wang M, Lu W, et al. Heterogeneous changes in mobility in response to the SARS-CoV-2 Omicron BA.2 outbreak in Shanghai. *Proc Natl Acad Sci U S A*. Oct 17, 2023;120(42):e2306710120. [FREE Full text] [doi: [10.1073/pnas.2306710120](https://doi.org/10.1073/pnas.2306710120)] [Medline: [37824525](https://pubmed.ncbi.nlm.nih.gov/37824525/)]

Abbreviations

SEIR: Susceptible-Exposed-Infected-Recovered

Edited by LCM Lau; submitted 05.12.23; peer-reviewed by Y Su, R Liu; comments to author 21.02.24; revised version received 31.03.24; accepted 19.04.24; published 28.06.24

Please cite as:

Liu C, Holme P, Lehmann S, Yang W, Lu X

Nonrepresentativeness of Human Mobility Data and its Impact on Modeling Dynamics of the COVID-19 Pandemic: Systematic Evaluation

JMIR Form Res 2024;8:e55013

URL: <https://formative.jmir.org/2024/1/e55013>

doi: [10.2196/55013](https://doi.org/10.2196/55013)

PMID:

©Chuchu Liu, Petter Holme, Sune Lehmann, Wenchuan Yang, Xin Lu. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 28.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.