A Methodological Critique of COVID-19 Studies Based on Cell-Phone Mobility Data

Shiv Yucel

1 Introduction

1.1 Background

In the context of the ongoing COVID-19 pandemic, understanding the mobility of individuals and populations is essential to model, predict, and analyse the spatio-temporal risk of disease spread (Oliver, Matic, & Frias-Martinez, 2015). The global adoption of mobile phones has granted researchers access to location-based data at a scale and scope that is a leap from household survey data previously available. By Schroeder's (2014) definition, this grants it the status of 'Big Data'. The spatial and temporal resolution of passively collected cell-phone mobility data has proven invaluable for research on human-transmitted diseases, which benefits from a continuous understanding of population flow and contact patterns (Oliver et al., 2015). While novel and exciting, the emergence of COVID-19 research using mobile phone location data is not without methodological limitations.

1.2 Design of Analysis

Using Amaya, Biemer, and Kinyon's (2020) Total Error Framework (TEF), this analysis will systematically critique current methodologies used to study COVID-19 based on cell-phone mobility data, and suggest possible improvements. The Total Error Framework is a general framework used to identify, describe, and interpret errors in Big Data sets, regardless of size or dynamism (Amaya et al., 2020). Big

Data research, in comparison to traditional survey-based research, often seeks to identify already existing data sets that can provide insights to a research question (Amaya et al., 2020; Salganik, 2019). As a Big Data source, cell-phone mobility data and its application to studying COVID-19 is a highly suitable topic to be analysed under the Total Error Framework. Key methodological issues when using cell phone data relate to the type of commercial data set used in the analysis. Rather than focus on a narrow subset of papers, this essay will cluster the COVID-19 literature around this data-type axis, providing a more complete perspective of existing methodologies. Prior to the methodological critique, this analysis will summarize the two widely used commercial mobile phone data types.

2 Mobile Phone Data Sources

2.1 Call Detail Records

Call detail records (CDRs) are event-driven data sources, providing a location and timestamp when a device uses a cell-phone network service (Oliver et al., 2015). Common actions which generate a CDR are sending and receiving calls, or short message services (SMS) (Wesolowski et al., 2014; Oliver et al., 2015). The location provided by the CDR is not a precise geo-located position. Instead, it is the cell-phone tower that provided the communication – their position around the cell tower remains unknown (Oliver et al., 2015). Mobile network logs are standardised and do not rely on algorithms that change over time, and thus are at low risk for algorithmic biases (Oliver et al., 2015). CDRs are generated by cell network companies for billing purposes and can be adapted for research purposes.

2.2 Mobile Phone Applications

Various companies have partnerships with mobile phone applications to harvest the locations of app visits to provide analytics into their user base. This data has

has been re-purposed for research projects given the severity of COVID-19 and the need to understand its spread (Grantz et al., 2020). Similar to CDRs, they are event-based, and only collect information when an individual is using a partnered app. As an example, In Loco, a Brazilian analytics company designed to help apps gain insights on their customer base, uses software development kits (SDKs) to anonymously track users' locations while using the app (Peixoto, Marcondes, Peixoto, & Oliva, 2020). In comparison with CDRs, application-based mobility data is more spatially granular and can provide specific geo-coordinates.

3 Coverage Error

Coverage error, as defined in the Total Error Framework, has three main components: undercoverage, overcoverage, and duplicative entries (Amaya et al., 2020). Undercoverage is the result of some target population units being excluded in the frame. It can lead to coverage bias if the magnitude is great enough, and the covered and uncovered populations differ in ways related to the object of measourement (Amaya et al., 2020). Overcoverage is the result of including units outside of the target population, or from duplication. Duplication occurs when the same unit is measured multiple times (Amaya et al., 2020). Overcoverage may introduce biases if the units outside the target population differ in ways related to the characteristic of interest, and duplicates may bias the data by overrepresenting some units in the estimates (Amaya et al., 2020).

3.1 Undercoverage in Commercial Data Sets

COVID-19 studies using cell-phone mobility data, whether CDR-or app-based, leverage a data source that was created for commercial rather than mobility research purposes. Coverage is determined by the market share of the data provider, without concern for creating a representative sample of the target population (Tizzoni et al., 2014). Despite mobile phones becoming globally ubiquitous, with adoption from nearly all socio-economic statuses, possession and use rates vary dramatically based on demographic and income statuses (Kraemer et al., 2020). The question of undercoverage becomes two-fold: Who possesses and uses phones regularly, and what subset is captured by the company's market share?

Current literature on the application of cell-phone mobility data for public health research posits that wealthier individuals, especially educated urban males, tend to possess and use phones more frequently than lower-income individuals (Kraemer et al., 2020). This undercoverage may intersect with existing racial and gender disparities (Kraemer et al., 2020; Oliver et al., 2020). Additionally, the elderly are frequently underrepresented in samples where mobile phone possession is a requirement (Grantz et al., 2020). If the movement of included groups differs from those whose movement is excluded or frequently undetected, then undercoverage bias is present within the sample (Aleta, Mart, Bakker, Pastore, & Ajelli, 2020).

App-based data, requiring advanced smartphone functionalities, may further exacerbate the income-related biases on cell-phone ownership and use (Oliver et al., 2015). The partnered apps generating the data may cater specifically to some subsets of the greater population. Different types of apps generate observations at varying frequencies, such as a mobile game versus a food delivery service – a separate source of unequal coverage. Finally, app-based data may require users to opt-in or opt-out of sharing location data, which introduces another source of undercoverage bias if specific population groups opt-in at different rates (Grantz et al., 2020).

Aside from who owns and uses cell-phones, the market share of a company may cater to specific individuals or areas (Grantz et al., 2020). Regions or groups with low coverage by the data provider may be excluded or underrepresented in the analysis, a problem for both CDR-and app-based data (Grantz et al., 2020). The distribution of app-based coverage across São Paulo in a single day is presented in Figure 1, from Peixoto et al. (2020). Large sample sizes in Big Data sets will not ameliorate systematic biases, and may in fact magnify them (Kaplan, Chambers, & Glasgow, 2014).

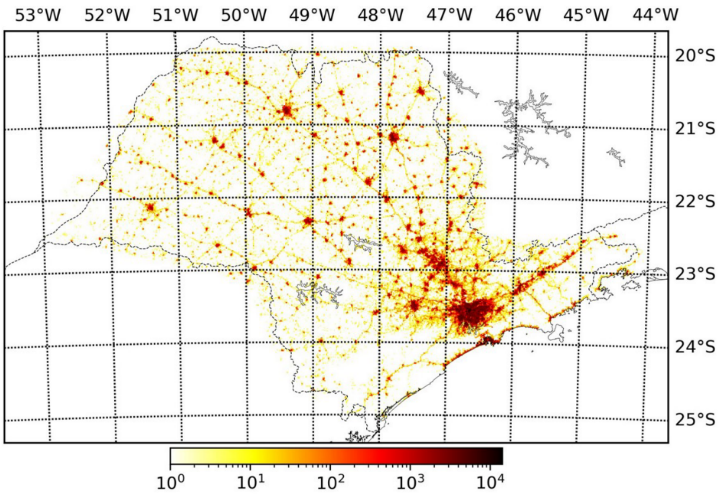


Figure 1: Distribution of App-Based Observations in São Paulo (colorbar represents density) (Peixoto et al., 2020)

3.2 Overcoverage and Duplicative Entries

Klein et al. (2020) use anonymised mobile phone location data to detect how commuting patterns in the US interact with the COVID-19 outbreak, and must infer home locations from frequently visited locations. Their study, using data from April and May 2020, may falsely include temporary visitors during this period as members of the regional population, introducing overcoverage error (Klein et al., 2020). Duplication error may occur when an individual uses multiple SIM cards or cell-phones, which neither CDR nor app-based mobile phone data can distinguish (Grantz et al., 2020). Wealthier individuals are those most likely to possess multiple SIM cards or phones, leading to duplication error risks among those wealthier in a sample (Grantz et al., 2020).

4 Sampling Error

Sampling error, as defined in the TEF, includes errors that occur as a consequence of analyzing a subset of the population, in lieu of the entire population of interest (Amaya et al., 2020).

4.1 Cell-Phone Users as Subset of True Population

Research on COVID-19 using CDR-or app-based data, which often aims to model trends in movement across entire populations, leverages large sample sizes from cell-phone data providers. Peixoto et al. (2020) conducted a study using In Loco's app-based mobility measures to evaluate how movement patterns affected the time it took for COVID-19 to infect each Brazilian city. This study has approximately 10 million unique users in the state of São Paulo. Despite their large sample size, this represents less than a quarter of São Paulo, which has a population of 48 million (Peixoto et al., 2020). Without control over the anonymised sample, the probability of a unit being included in the sample cannot be determined by researchers. This can be particularly problematic for cell phone data research which often excludes individuals under 18, excluding a major subset of the younger population which is of great concern to epidemiological studies (Flasche & Edmunds, 2021).

5 Specification Error

Specification error is defined as the failure of the items being measured to capture the concept required by the research question (Amaya et al., 2020). When re-purposing commercial mobility data sets for COVID-19 research, this error type is present for both mobility-based and epidemiological concepts.

5.1 Operationalising Mobility

Timestamped locations for an individual are operationalised as true measures of a population's movement, yet only capture the fraction of movements where they use their phone for cell-services or a partnered app, depending on the data source. In Vinceti et al.'s (2020) analysis of the efficacy of lockdown in Italy with CDR data, a journey was determined as 'finished' if the individual did not produce any new observations for the next hour. If individuals continued to travel without using their phones, this would be excluded in the data set.

5.2 Operationalising Disease Risk

From an epidemiological perspective, it is not only movement that aims to be accurately modeled, but the risk of COVID-19 disease spread, which cannot be completely captured with cell-phone mobility data. To capture the spread of risk, current literature often employs compartmental epidemiological models. These models, also known as SIR (susceptible, infectious, and recovered) models, assign individuals the status of susceptible, infectious, and recovered, represented by S_t, I_t, R_t in the following equation (Aleta et al., 2020). The population, N_j, is modeled over time by:

$$N_j = S_j(t) + I_j(t) + R_j(t) \tag{1}$$

The population remains constant, but the share of individuals in the SIR states vary over time depending on the transmission of disease. This model, previously used with traditional survey-based data, has been adapted to cell-phone mobility models (Colizza, Barrat, Barthelemy, & Vespignani, 2006; Aleta et al., 2020; Peixoto et al., 2020). The probability of a susceptible individual becoming infected at time period t is modeled by βdt, where β represents a transmission parameter defined by the reproductive number of the disease (Colizza et al., 2006). Infected individuals recover with a probability of udt, where u is the average disease duration. Once recovered, they are unable to re-enter any other state. Individuals are dynamically assigned across time based on their frequency and length of exposure to different individuals and locations (Colizza et al., 2006).

Despite large sample sizes of high spatial and temporal resolution, cell-phone mobility data is an incomplete data source to capture the concept of disease risk with SIR models. Neither data source is able to distinguish the mode of transportation, which strongly influences an individual's probability of infection, such as whether one uses a private car or public train (Barbieri et al., 2021). As a result, studies including Chang et al.'s (2020) that analyse the effects of mobility on virus spread with a sample size of 98 million, focus only on visitation destination and home locations – excluding transport type. The 552,758 destinations studied by Chang et al. (2020) include restaurants, grocery stores, and malls. Studies employing this research design are unable to truly capture disease risk, excluding one's occupation, whether they wear masks, or have access to healthcare, among other crucial details (Chang et al., 2020; Gatalo, Tseng, Hamilton, Lin, & Klein, 2020). Cross-relating external social information to the data set may improve these issues, further discussed in Section 1.9 on processing errors.

A final issue is that individuals who recover once from COVID-19 are at risk of re-infection, therefore the finality of a 'recovered' individual in the SIR model does not reflect one's risk in reality (Iwasaki, 2021). Researchers on COVID-19 using cell-phone mobility data would do well to qualify their terminology and explicitly state these underlying assumptions.

6 Nonresponse/Missing Data Error

The fourth section in the TEF, often confounded with undercoverage, is nonresponse or missing data error (Amaya et al., 2020). With cell-phone mobility data, this can result from a lack of data for analysis for an individual in the sample or anonymisation measures from the data provider.

6.1 Insufficient Data for Analysis

For individuals within the sample of mobile phone users, missing data corresponds to individuals who do not generate adequate data for analysis (Amaya et al., 2020). When using cell-phone mobility data as a proxy for mobility in epidemiological studies, researchers require a minimum amount of data to estimate commuting patterns (Tizzoni et al., 2014). As an example, Peixoto et al. (2020) required at least two movements in a 24 hour period to include an individual in the population flow.

6.2 Erasure of Data for Anonymisation Purposes

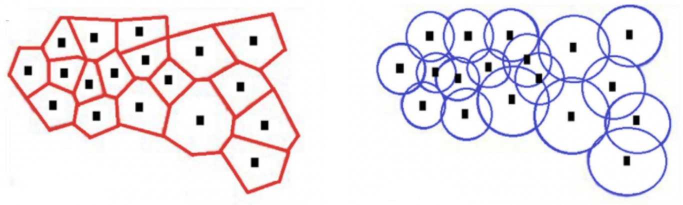
Missing data may additionally be created due to privacy concerns from the data provider. In a study of human mobility dynamics and their interaction with non-pharmaceutical interventions, Kang et al. (2020) compare the origin census blocks of visitors to places of interest. If only one individual visits a place of interest, the data provider erases the visit to protect the privacy of the user – generating missing data for that user in the sample (Kang et al., 2020). Coverage error may interact with missing data error if users with insufficient observations vary systematically in mobility patterns compared to those without (Amaya et al., 2020).

7 Measurement/Content Error

Measurement and content error arise from processes that introduce noise and decrease estimator precision, occurring from both mechanical and anonymity measures by cell-phone data providers (Amaya et al., 2020).

7.1 Noise in CDR Data

When using CDR data to estimate an individual's location, the cell tower providing the service is returned. To generate areas where an individual may be when signaling a cell tower, a Voronoi tesselation is often used to map the individual to a distinct unit. A Voronoi tesselation divides a region into cells around each tower, where each cell corresponds to the region closest to that tower, shown in Figure 2a. In reality, the coverage of cell-phone towers overlap (Figure 1b), and the tesselation area may not correspond to the location of the individual signaling the tower (Oliver et al., 2015).



(a) (b)

Figure 2: Tesselation (a) versus Cell Tower Coverage (b), from Oliver et al. (2015)

7.2 Noise from Anonymisation Practices

Additional measurement error may be introduced from anonymisation practices by data providers. In Kang et al.'s study of visits to places of interest after non-pharmaceutical interventions, the app-based data records places of interest with between 2 and 4 visitors as having 4 visitors, to protect differential privacy. This introduces an inflationary bias for low numbers of visitors, creating a measurement error. In this study, only when there are 5 or more visitors to a region is the true number of visits to the location included (Kang et al., 2020).

8 Processing Error

The Total Error Framework defines processing errors as those stemming from data entry, coding, editing, and variable transformations or conversions (Amaya et al., 2020). These most often arise when adding social elements to a mobility-based COVID-19 research question, through linking existing data sets.

8.1 Geographic Linking of Data Sets

Both CDR-and app-based geo-location data are 'shallow' and only provide locations of varying granularity across different time periods. To add depth to the mobility data, researchers may map individuals to census regions to incorporate racial or demographic characteristics into the sample. The process of mapping individuals to census regions may introduce errors due to imprecise linkage with existing records, a problem common to Big Data (Amaya et al., 2020).

When linking mobility data to census regions, spatial aggregation or de-aggregation must occur to geographically link the two data sources (Tizzoni et al., 2014). A study by Li, Sabrina, Pereira, Prete, Zarebski, and Emanuel uses a hexagonal social isolation index generated by app-based data to determine social and racial inequalities in COVID-19 risk. Each grid corresponds to the percentage of mobile phone users who left their region during a given day. To map the hexagons to census regions for racial and income-based inferences, Li, Sabrina, Pereira et al. (2020) employ a technique known as dasymetric interpolation. This method passes census data to a grid of 200 m^2, then re-aggregates them to match the hexagonal regions using an aerial intersection. While allowing researchers to infer information for regions that lie between census blocks, it assumes a homogeneous distribution of racial and income-based proportions within census blocks – a strong assumption that is prone to error during geographic matching.

8.2 Outdated Data Linkage

Without the ability to gather real-time data on the traits of individuals in the mobile phone dataset, researchers often assume that racial and income level proportions have remained constant since the last census. The risks of time-lagged data vary based on the proximity of the most recent census to the period of interest. In their study of behavior influencing the second wave of COVID-19 in the United States, Aleta et al. (2020) match individuals to household sizes based on the 2018 US census. In contrast, Li, Sabrina, Pereira et al. (2020) match the 2020 mobility data to 2010 Brazilian census tract data, the most recently available demographic data-set. This demographic characteristic mapping assumes that proportions have not changed in the decade prior to the study, a potential source of processing error. Big Data and survey research should not be mutually exclusive, and researchers facing severe time lags would do well to design their own surveys across census regions to gain a more proximate understanding of their populations of study.

9 Modeling/Estimation Error

Modeling and Estimation errors are the product of deficiencies in researchers' abilities to adjust for missing data or coverage errors with weighting, or to impute missing data. Without an understanding of the missing data mechanisms in cell-phone mobility data, this is a great challenge (Amaya et al., 2020).

9.1 Insufficient Information for Weighting

Cell-phone mobility data faces the same challenges as other Big Data sources for modeling and estimation error, due to a lack of knowledge of the forces that lead to individuals being included and excluded in the sample (Amaya et al., 2020). With cell-phone mobility data, issues of undercoverage and missing data are compounded by the anonymisation of data for privacy purposes, and the lack of goldstandard data to validate a sample of mobile phone users (Grantz et al., 2020). Post-stratification sampling, a method to create representative samples using weighting after the collection of data, is infeasible without user metadata and an understanding of the true distribution of users (Kolenikov, 2016).

10 Analytic Error

Analytic errors in the Total Error Framework refer to errors made by researchers and clients in interpreting the findings of research, mainly found when interpreting the results of SIR models (Amaya et al., 2020).

10.1 Ambiguity in SIR Models

Analytical errors in COVID-19 studies using cell-phone mobility data arise from a lack of discussion regarding the assumptions of SIR models. While susceptibility depends on one's movement patterns and exposure to other individuals, it assumes that individuals at a given destination mix with an equal probability of contact (Tolles & Luong, 2020). This fails to consider the dynamics of real networks which are more limited and selective in scope (Tolles & Luong, 2020). Another source of analytic error is the ambiguity of the term 'recovered'. It refers to individuals who are no longer contagious, *not* necessarily individuals who are no longer ill – an important distinction given the long term health effects of COVID-19 (Tolles & Luong, 2020). This may underestimate individuals still suffering from the disease, leading to a misinterpretation of the overall risks to certain populations. Figure 3 represents risk calculations from Peixoto et al. (2020) across regions in Rio de Janeiro based on an SIR model. Further, when an individual dies of COVID-19 they are assigned to the 'recovered' group, leading to an ambiguous interpretation of a population's 'recovery rate'. To improve analytic errors in COVID-19 studies using cell-phone mobility data, researchers should be cautious about their terminology and thoroughly explain the assumptions embedded in the results of their model.

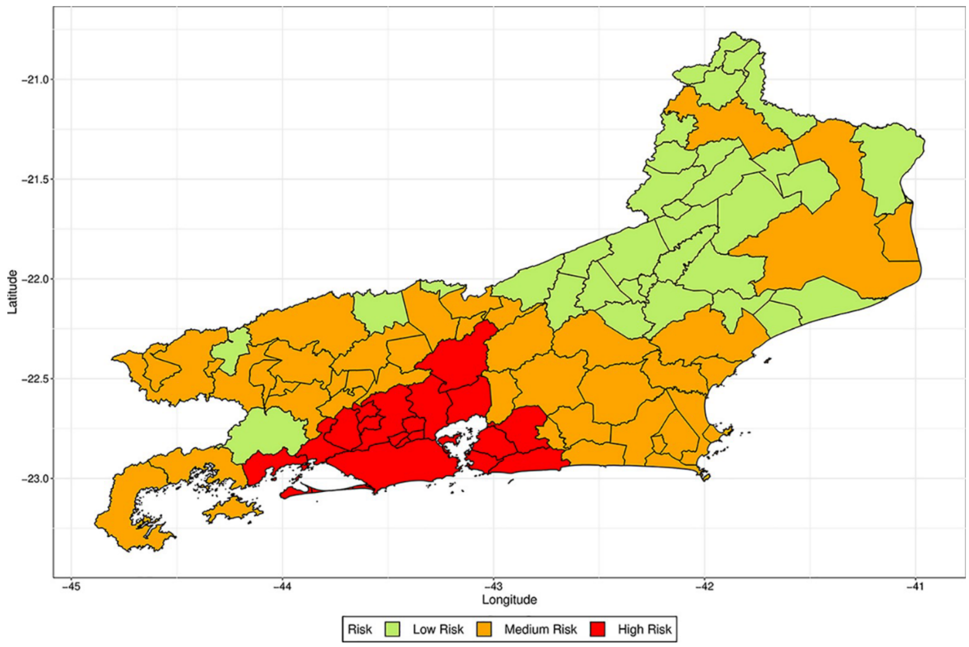


Figure 3: Labelled Risk Based on SIR Model in Rio de Janeiro (Peixoto et al., 2020)

11 Conclusion

Using Amaya et al.'s Total Error Framework to systematically tackle 8 sources of error, it is evident that COVID-19 research using cell-phone mobility data, while promising, has significant room for improvement. A dominant source of such errors is the inability of researchers to control the sample selection process and understand their traits, due to a reliance on anonymised data from private companies. The SIR model, while a useful tool for modeling disease with mobile phone data, is limited in its ability to capture the concept of 'risk', partially due to a lack of deeper data on individuals in the sample, such as occupation. The primary take-away from this section of the essay is that Big Data opens the door to previously impossible research on the study of disease but should not be exempt from critical analyses, especially in the context of a global emergency. Understanding the limitations of such novel methodologies is an important step to increase the robustness of future research on the current pandemic, and to prepare for the next.

References

Aleta, A., Mart, D., Bakker, M. A., Pastore, A., & Ajelli, M. (2020). Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas. , 1–18.

Amaya, A., Biemer, P. P., & Kinyon, D. (2020). Total Error in a Big Data World: Adapting the TSE Framework to Big Data. *Journal of Survey Statistics and Methodology*, *8*(1), 89–119. Retrieved from `https://doi.org/10.1093/jssam/smz056` doi: 10.1093/jssam/smz056

Barbieri, D. M., Lou, B., Passavanti, M., Hui, C., Hoff, I., Lessa, D. A., ... Rashidi, T. H. (2021). Impact of COVID-19 pandemic on mobility in ten countries and associated perceived risk for all transport modes.(Research Article)(Report). *PLoS ONE*, *16*(2), e0245886.

Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2020). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, *589*(January). Retrieved from `http://dx.doi.org/10.1038/s41586-020-2923-3` doi: 10.1038/s41586-020-2923-3

Colizza, V., Barrat, A., Barthelemy, M., & Vespignani, A. (2006). The role of the airline transportation network in the. *Pnas*, *103*(7), 2015–2020.

Flasche, S., & Edmunds, W. J. (2021). The role of schools and school-aged children in SARS-CoV-2 transmission. *The Lancet Infectious Diseases*, *21*(3), 298–299. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1473309920309270` doi: https://doi.org/10.1016/S1473-3099(20)30927-0

Gatalo, O., Tseng, K., Hamilton, A., Lin, G., & Klein, E. (2020). Associations between phone mobility data and COVID-19 cases. *The Lancet Infectious Diseases*. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1473309920307258` doi: https://doi.org/10.1016/S1473-3099(20)30725-8

Grantz, K. H., Meredith, H. R., Cummings, D. A., Metcalf, C. J. E., Grenfell, B. T., Giles, J. R., ... Wesolowski, A. (2020). The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature Communications*, *11*(1), 1–8. Retrieved from `http://dx.doi.org/10.1038/s41467-020-18190-5` doi: 10.1038/s41467-020-18190-5

Iwasaki, A. (2021). What reinfections mean for COVID-19. *The Lancet Infectious Diseases*, *21*(1), 3–5. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1473309920307830` doi: https://doi.org/10.1016/S1473-3099(20)30783-0

Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., & Kruse, J. (2020). Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. *Scientific Data*, *7*(1), 1–13. doi: 10.1038/s41597-020-00734-5

Kaplan, R. M., Chambers, D. A., & Glasgow, R. E. (2014, aug). Big data and large sample size: a cautionary note on the potential for bias. *Clinical and translational science*, *7*(4), 342–346. Retrieved from https://pubmed.ncbi.nlm.nih.gov/25043853https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5439816/ doi: 10.1111/cts.12178

Klein, B., LaRock, T., McCabe, S., Torres, L., Friedland, L., Privitera, F., ... Chinazzi, M. (2020). Reshaping a nation: Mobility, commuting, and contact patterns during the COVID-19 outbreak..

Kolenikov, S. (2016). Post-stratification or non-response adjustment? *Survey Practice*, *9*(3). doi: 10.29115/SP-2016-0014

Kraemer, M. U., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., ... Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497. doi: 10.1126/science.abb4218

Li, Sabrina, Pereira, R., Prete, C., Zarebski, A., & Emanuel, L. (2020). Social and racial inequalities in COVID-19 risk of hospitalisation and death across São Paulo state, Brazil. , 1–23.

Oliver, N., Matic, A., & Frias-Martinez, E. (2015). Mobile Network Data for Public Health: Opportunities and Challenges. *Frontiers in Public Health*, *3*(August), 1–12. doi: 10.3389/fpubh.2015.00189

Oliver, N., Oliver, N., Lepri, B., Lepri, B., Sterly, H., Lambiotte, R., ... Vinck, P. (2020). Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*, *6*(23), 1–7. doi: 10.1126/sciadv.abc0764

Peixoto, P. S., Marcondes, D., Peixoto, C., & Oliva, S. M. (2020). Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. *PLoS ONE*, *15*(7 July), 1–23. doi: 10.1371/journal.pone.0235732

Salganik, M. J. (2019). *Bit by bit : social research in the digital age*. Princeton ; Oxford.

Schroeder, R. (2014). Big Data and the brave new world of social media research. *Big Data and Society*, *1*(2), 2053951714563194. Retrieved from https://doi.org/10.1177/2053951714563194 doi: 10.1177/2053951714563194

Tizzoni, M., Bajardi, P., Decuyper, A., Kon Kam King, G., Schneider, C. M., Blondel, V., ... Colizza, V. (2014). On the Use of Human Mobility Proxies for Modeling Epidemics. *PLoS Computational Biology*, *10*(7). doi: 10.1371/journal.pcbi.1003716

Tolles, J., & Luong, T. (2020). Modeling Epidemics With Compartmental Models. *JAMA*, *323*(24), 2515–2516. Retrieved from https://doi.org/10.1001/jama.2020.8420 doi: 10.1001/jama.2020.8420

Vinceti, M., Filippini, T., Rothman, K. J., Ferrari, F., Goffi, A., Maffeis, G., & Orsini, N. (2020). Lockdown timing and efficacy in controlling COVID-19 using mobile phone tracking. *EClinicalMedicine*, *25*. doi: 10.1016/j.eclinm.2020.100457

Wesolowski, A., Stresman, G., Eagle, N., Stevenson, J., Owaga, C., Marube, E., ... Buckee, C. O. (2014). Quantifying travel behavior for infectious disease research: A comparison of data from surveys and mobile phones. *Scientific Reports*, *4*. doi: 10.1038/srep05678