

REVIEW

Open Access



Trending on the use of Google mobility data in COVID-19 mathematical models

Yang Deng^{1,2}, Hefei Lin¹, Daihai He^{1*}  and Yi Zhao^{2*}

*Correspondence:

daihai.he@polyu.edu.hk;
zhao.yi@hit.edu.cn

¹Department of Applied
Mathematics, Hong Kong
Polytechnic University, Hong Kong,
China

²School of Science, Harbin Institute
of Technology, Shenzhen, 518055,
China

Abstract

Google mobility data has been widely used in COVID-19 mathematical modeling to understand disease transmission dynamics. This review examines the extensive literature on the use of Google mobility data in COVID-19 mathematical modeling. We mainly focus on over a dozen influential studies using Google mobility data in COVID-19 mathematical modeling, including compartmental and metapopulation models. Google mobility data provides valuable insights into mobility changes and interventions. However, challenges persist in fully elucidating transmission dynamics over time, modeling longer time series and accounting for individual-level correlations in mobility patterns, urging the incorporation of diverse datasets for modeling in the post-COVID-19 landscape.

Keywords: Mathematical models; Google mobility data; Contact matrix; Transmission rate; The basic reproduction number; COVID-19

1 Introduction

Due to the highly contagious nature of COVID-19, reducing social interactions and community movement has been crucial in lowering transmission rates [21, 30, 32]. Interventions have been adopted to reduce the transmission of COVID-19, including the practising of social distancing, self-isolation or quarantine [15, 17], and so on. The implementation of interventions in response to infectious disease outbreaks is not new, and these methods aiming to reduce social contact and limit mobility have been used for centuries, as adopted in the outbreak of MERS and SARS epidemics [6, 37]. However, despite the historical knowledge of the link between mobility and disease, quantifying this relationship in detail has been challenging, especially over large geographical areas and for large populations. In response to the COVID-19 pandemic, academic researchers have dedicated significant efforts to study the connection between human mobility and COVID-19 transmission. They have utilized various datasets and mathematical models in different countries and regions [14, 26, 27]. An example of such datasets is provided by Google. Google released data collected from users accessing its applications through handheld devices. The “Community Mobility Reports” (CMR) [1] from Google showcase alterations in activity and mobility across various location types, comparing the period before the global spread of COVID-19. Given the lack of alternative global data sources for these factors, Google mobility data serves as a reliable indicator of the impact that health recommenda-

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

tions and government restrictions have had on social activity and movement. It provides distinctive and valuable insights into changes in mobility, presenting a unique opportunity to explore the correlation between mobility and disease incidence. Thus, researchers are progressively exploring methods to integrate Google mobility trends into COVID-19 research. Searching on PubMed with the terms “Google Mobility Data” and “COVID-19” generates over 288 results, while on Google Scholar, there are more than 694,000 matches for the same query.

In this review, we have delved into the extensive body of literature addressing using Google mobility data during the COVID-19 crisis. While these papers employ both statistical methods and mathematical models, including compartmental and metapopulation models, our primary focus centers on the utilization of Google mobility data in the context of COVID-19 mathematical modeling. We examined existing models incorporating Google mobility data in general and highlighted the use and effectiveness of Google mobility data to enhance traditional infectious disease models and discuss challenges that may arise with its burgeoning addition to the infectious disease modeling suite. We also discussed papers that did not employ the Google mobility data directly in the model but instead used it to validate the model performance or model input data source.

This paper is organized as follows: Sect. 2 describes the method of article collection; Sect. 3 details the application of Google mobility data in different aspects of epidemic modeling. Last but not least, some challenges, observations, and conclusions are summarized in Sect. 4.

2 Article collection

We initiated our search by exploring PubMed and Google Scholar for articles published between January 2020 and May 2023, aiming to encompass the latest research on the utilization of Google mobility data in COVID-19 models. The searching terms we used are “COVID-19”, “novel coronaviruses”, “2019-nCov”, “SARS-CoV-2”, “Google mobility data”, and “mathematical modeling”. This effort yielded around 100 relevant articles for our study. The selected articles apply various mathematical models to analyze, simulate, and predict the association between human mobility and COVID-19. We categorized these articles mostly into two distinct groups: those utilizing Google mobility data directly in the modeling process and those that use Google mobility data as reference to validate the model. The varieties of models employed in these papers are depicted in Fig. 1. This concentration led us to conduct a thorough examination of more than a dozen highly influential research studies.

3 Epidemic models

Upon reviewing these studies, we observed that they could be categorized based on the mathematical models utilized and whether they incorporate Google mobility data into the modeling framework. Therefore, in Table 1, we first categorized the most of collected articles into two groups, which modeled the COVID-19 dynamics with either compartmental or metapopulation models. Subsequently, in Sects. 3.1 and 3.2, we delve deeper into this classification, specifically considering whether Google mobility data is incorporated into their modeling procedures.

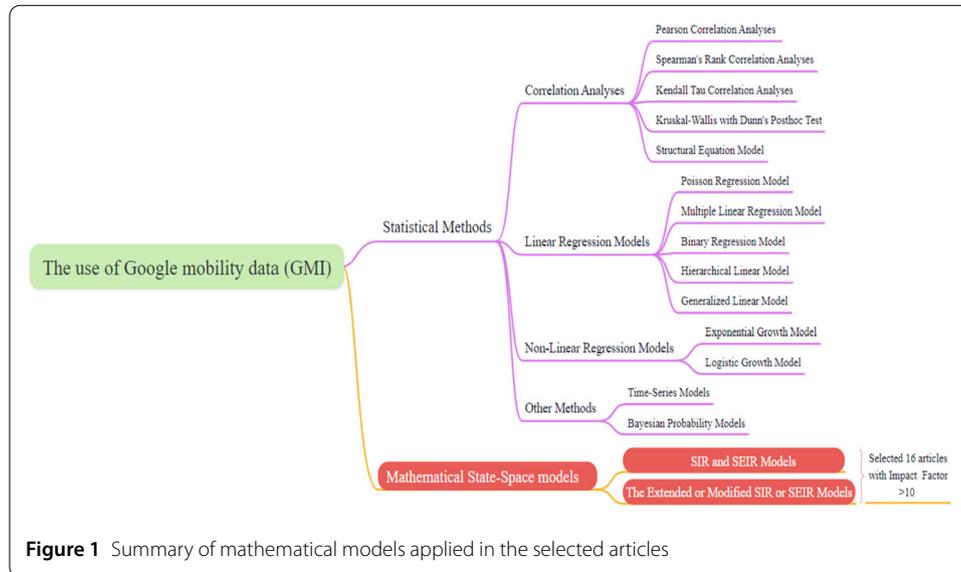


Figure 1 Summary of mathematical models applied in the selected articles

Table 1 Summary of the epidemic models: the first column describes the main categories; the second column shows the sub-categories; the last column presents the countries studied in each references

Category	Subcategory	The countries studied in each reference	
Compartmental models	SIR-type	US [34]	
	Age structured SIR-type	Illinois, USA [38]	
	SEIR-type	US [8]	
		Connecticut [9]	
Compartmental models	Age structured SEIR-type	Kenya [4]	
		US [18]	
		UK, South Africa, Brazil [39]	
	Age structured SEIR-type	Philippines [9]	
		Israel [13]	
		Ontario, Canada [19]	
Metapopulation models	Age structured	France [26]	
		England and Wales [35]	
		Age structured	Santiago de Chile [14]
		Age unstructured	China [27]
		Age unstructured	Europe [28]
		US [7]	

3.1 Models that incorporated the Google mobility data into the epidemic model

As shown in Table 1, the focus of this review is on compartmental models and metapopulation models. Compartmental models consider a single population divided according to health statuses and, in some cases, age structure. In fact, it is extremely common to describe individuals' progression in the different phases of a disease (i.e., natural history of the disease) via compartments, each one representing health statuses. Susceptible, asymptomatic infectious, infectious, hospitalized, and recovered statuses are classic examples. All compartmental models in the papers above are variations of two basic archetypes: the susceptible-infectious-recovered (SIR) models or the susceptible-exposed-infectious-recovered (SEIR) models. We first reviewed the modeling work based on SIR and SEIR models.

Several articles have incorporated Google mobility data directly into modeling COVID-19 dynamics through methods such as simplifying and expressing via effective reproduc-

tion functions, contact matrices, and rescaling key parameters. Unwin et al. [34] used a Bayesian hierarchical semi-mechanistic model of COVID-19 transmission in the states of US, accounting for nonpharmaceutical interventions (NPIs) and mobility at a state level. Google mobility data was used in the parametric definition of R_t , the time-dependent reproduction number:

$$R_{t,m} = R_{0,m} \cdot f \left(- \left(\sum_{k=1}^2 X_{t,m,k} \alpha_k \right) - \sum_{l=1}^2 Y_{t,m,l} \alpha_{r(m),l}^{region} - Z_{t,m} \alpha_m^{state} - \epsilon_{m,w_m(t)} \right),$$

where $f(x) = \frac{2 \exp(x)}{1 + \exp(x)}$; $X_{t,m,k}$ are the covariates that exert the same effect for all states; $Y_{t,m,l}$ are region-specific effects; $r(m) \in \{1, \dots, R\}$ represents the region that the state is in; $Z_{t,m}$ is a covariate that has a state-specific effect, and $\epsilon_{m,w_m(t)}$ is modeled as a weekly autoregressive AR(2) process centred around 0. The covariates selected are $X_{t,m,1} = M_{t,m}^{average}$, $X_{t,m,2} = M_{t,m}^{residential}$, $Y_{t,m,1} = 1$ (an intercept), $Y_{t,m,2} = M_{t,m}^{average}$, and $Z_{t,m} = M_{t,m}^{average} \cdot M_{t,m}^{average}$ is an average of variables for retail and recreation, groceries and pharmacies, and workplaces. $M_{t,m}^{residential}$ is the variable for places of residences.

Similarly, another group of researchers [8] also developed an SEIR-type compartmental model to evaluate the impact on the COVID-19 epidemic in each state of the United States via incorporating mobility data, confirmed case data and contact tracing. To estimate contact rates, the authors employed several types of mobility data (i.e., Unacast [33], Google [1], OpenTable [31]). Within the model, the influence of social distancing, hygiene measures, and reopening is characterized by a time dependence of the contact rate $c(t)$: $c(t) = c_0 \times [\theta(t) + (1 - \theta_{min}) \times r(t)]$ and the probability of transmission per infected contact β : $\beta(t) = \beta_0 \times \theta(t)^n$. Several mobility data are applied to fit the contact rate model $c(t)$, aiming to derive the prior distributions for parameters. The authors found that Google’s “retail and recreation” ($\gamma^2 = 0.49$) and Unacast ($\gamma^2 = 0.52$) generate the highest R-squared values. In summary, the findings indicate the necessity to broaden the utilization of mobility data sources for constructing prior distributions, as opposed to merely incorporating such data directly into modeling contact rates.

In another paper published in 2021, authors [18] present a deterministic SEIR compartmental framework to forecast severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections and evaluate the effects of nonpharmaceutical interventions within the United States, including analysis at the state level as well. The duration of the research period was longer than that of the previous two studies. Interestingly, in their model, not only β is considered a function of time, but the force of infection is modulated by a mixing parameter α defined in such a way that $\lambda(t) = \beta(t)(I_1 + I_2)^\alpha / N$, where I_1 and I_2 describe pre-symptomatic and symptomatic individuals. They use four data sources on human mobility to construct a composite mobility indicator by a linear regression model, linking the implementation of different NPIs. Those sources include not only Google Community Mobility reports [1] but also Facebook Data for Good [12], SafeGraph [29], and Descartes Laboratories [11]. For Google mobility data, they take the average of the percentage change in the “Retail and recreation,” “Transit stations,” and “Workplaces” to represent the mobility trend most strongly affected by the social distancing measures. Research results confirm the effectiveness of NPIs under different scenarios.

Other authors [4] also managed to construct a modification of the SEIR scheme using data from another country, Kenya, with a compartment W to account for the portion of

recovered individuals that return from a completely protected state to a partially protected state due to waning immunity. The authors model the SARS-CoV-2 dynamics in each of the 47 Kenyan counties as a two-group SEIRW transmission process with differences in their abilities to reduce social mobility. The per capita forces of infection on individuals in the two groups, lower and higher social-economic groups, denoted respectively $\lambda_L(t)$ and $\lambda_U(t)$, are described as follows:

$$\lambda_L(t) = \frac{\gamma R_0(t)}{N_L} (\varepsilon c_L(t) I_L(t) + (1 - \varepsilon) c_U(t) I_U(t)),$$

$$\lambda_U(t) = \frac{\gamma R_0(t)}{N_U} ((1 - \varepsilon) c_L(t) I_L(t) + \varepsilon c_U(t) I_U(t)).$$

The estimation of the proportion $c_U(t)$ of the higher socioeconomic group interacting in locations outside the home is determined through the average change in the “retail and recreation”, “grocery and pharmacy”, “transit stations”, and “workplaces” settings in Google mobility trend data. The authors posit that Google mobility data is more effective in depicting access trends for the higher socioeconomic group when visiting locations outside the home, attributed to their ownership of smartphones.

Based on case and mortality data, Yang and Shaman [39] introduced an SEIR-type for estimating the epidemiological characteristics of emerging SARS-CoV-2 variants. In a more detailed manner, authors utilize climate data to gauge the seasonality of diseases, mobility data to illustrate the comprehensive effects of non-pharmaceutical interventions (NPIs), and vaccination data to consider alterations in population susceptibility resulting from vaccination efforts, all within the contexts of the United Kingdom, South Africa, and Brazil. The mobility data derived from Google Community Mobility Reports, aggregate the relative mobility as observed in “Retail and recreational”, “Transit stations”, and “Workplaces” as the function m_t and the estimated seasonal trend then are used to adjust the transmission rate β_t . Their research results indicate that the NPIs can suppress the rise of the B.1.1.7, B.1.351, P.1, and the wild-type variants, and the continued NPIs will reduce infection resurgence. Based on the aforementioned papers, it can be inferred that diverse forms of mobility data have been incorporated into the SEIR model in varying ways. Even though the papers listed above achieved robust results, we would like to point out that articles listed above have a similarity in not considering an important element: the age-structure of the population. To reflect the effects of the nonpharmaceutical interventions in different age groups, several articles adopted SEIR-like models with age-structures to study the impact of different NPIs in different countries. Table 2 intuitively reflects the use and selection of mobility data for an age unstructured SEIR-type model.

The mentioned articles overlook a crucial factor: the age structure of the population. To capture the effects of nonpharmaceutical interventions across various age groups, several articles have employed SEIR-like models incorporating age structures to investigate the influence of different NPIs in various countries.

Caldwell et al. [5] used a modified age structured SEIR model with splitting the exposed and infectious groups into two sequential subgroups (SEPILR) to research the COVID-19 dynamics in Philippines. The Google mobility data was used to dynamically adjust the

Table 2 The use and selection of mobility data for an age unstructured SEIR-type model

Authors	The use of Google mobility data	Selection of mobility data
[8]	Contact rate	Unacast Dataset, OpenTable Dataset, six categories of Google mobility data
[4]	Force of Infection	“Retail and recreation”, “Grocery and pharmacy”, “Transit stations”, “Workplaces”
[18]	Scaling Transmission Rate	Facebook Dataset, SafeGraph, Descartes Labs Dataset, “Retail and recreation”, “Transit stations”, “Workplaces”
[39]	Scaling Transmission Rate	“Retail and recreation”, “Transit stations”, “Workplaces”

contact matrix. The force of infection was defined as follows:

$$\lambda_a(t) = \beta \left[\sum_{j,c} \frac{\varepsilon \times P_j}{N_j} \times C_{a,j}(t) + \sum_{j,c} \frac{I_{j,c} \times \iota_c + L_{j,c} \times \kappa_c}{N_j} \times C_{a,j}(t) \right],$$

where a is age, j and c are population groups, and P , I , and L are infectious groups. The contact matrix in a specific age group is adjusted as

$$C_t = h(t)C_H + s(t)C_S + w(t)C_W + l(t)C_L,$$

where C_H , C_S , C_W , and C_L are the age-specific contact matrices associated with households, schools, workplaces, and other locations [24]. $h(t)$ is a constant; $s(t)$ depends on the percentage of students attending educational institutions; $w(t)$ is a polynomial spline fitted to the Google mobility’s “workplaces” data; $l(t)$ is a polynomial spline fitted to the average Google mobility’s “retail and recreation” data, “grocery and pharmacy”, “parks”, and “transit stations”.

Jentsch et al. [19] developed an age-structured *SEPAIR* (susceptible, exposed, presymptomatic, asymptomatic, symptomatic, removed) model with 16 age classes to project COVID-19 mortality under four different COVID-19 vaccine scenerios in Ontario, Canada. The model takes population adherence to NPIs, changes to mobility patterns, and seasonality into consideration. The force of infection in the model can be modulated as follows:

$$\lambda_i(t) = \gamma \left[1 + s \sin\left(\frac{2\pi}{365}(t - \vartheta) - \frac{\pi}{2}\right) \right] \sum_{j=1}^{16} C_{ij}(t) \left(\frac{I_{sj} + I_{aj} + P_j}{N_j} \right),$$

where γ is the probability of transmission per contact, s represents seasonality, and ϑ is a seasonality phase. $C_{ij}(t)$ is the average number of contacts per day at workplaces, schools, households, and other locations, which can be represented as

$$C_{ij}(t) = C_{ij}^W(t) + C_{ij}^S(t) + (1 - \varepsilon_P x) \left(\overline{C}_{ij}^o + \overline{C}_{ij}^H \right),$$

where $C_{ij}(t)$ varies depending on individual adherence to NPIs as well as government shutdown policies. The authors utilize deviations from the baseline time spent at retail and recreational venues to signify population compliance with nonpharmaceutical interventions (NPIs). Consequently, the proportion $x(t)$ of individuals adhering to NPIs is determined by fitting the reduction in “Retail and Recreation” from Google mobility data.

The authors also fit a step function $f(t) = \varepsilon_W(\tanh k_1(t - t_{close}^W) - \tanh k_2(t - t_{close}^W))$ to the “Workplaces” field of the Google mobility data so as to obtain the values of ε_W, k_1, k_2 , thus revealing the workplace function $C_{ij}^W(t)$ and the school function $C_{ij}^S(t)$. Matrices \overline{C}_{ij}^O and \overline{C}_{ij}^H are merged under the assumption that NPIs in home has the same efficacy as other locations.

Pullano et al. [26] described the impact of age-specific contact activity in COVID-19 transmission in French regions with a stochastic discrete age-stratified SEIR structure. The authors adopted the social contact matrices measured in a survey in France in 2012 [3] as the baseline conditions for their model. The contact matrix incorporates both the nature of the activity and the location of contacts (such as home, school, workplace, etc.). Adjustments to the contact matrices serve as the basis for modeling intervention strategies. The application of the Google mobility data is to estimate the percentage change of individuals at the workplace to account for the work contact pattern changes. Moreover, in the analysis of model selection, the authors illustrate that accounting for changes in contact patterns during the exit phase of intervention measures provides a more accurate description of the epidemic trajectory.

Across the sea, with the data from England and Wales, Waterlow et al. [35] created a deterministic compartmental transmission model to examine the impact of cross-protection from seasonal coronavirus (HCoVs) on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Both seasonal HCoVs and SARS-CoV-2 have populations grouped in either susceptible (S), exposed (E), infectious (I), or recovered (R) compartments with five age groups. Due to the nonpharmaceutical interventions implemented, the authors split the contact matrices into three categories: school contact matrix, household contact matrix, and other contact matrix (originally from contacts in all other categories reported in the POLYMOD study [23]). Based on Google mobility data, the authors adjust the ‘other’ contact matrix with the average change in “retail and recreation”, “workplace”, “grocery and pharmacy”, and “transit stations” reported in the Google Community Mobility Reports. Table 3 intuitively reflects the use and selection of mobility data for age structured SEIR-type model.

Later in 2022, Gavish et al. [13] used a mathematical model that accounts for the age-stratification, vaccination, and booster administration, and waning immunity afterwards, to assess the population-level impact of the booster campaign in Israel. The social contact matrix used to model the infection process is composed of a time-varying linear combination of contact matrices. Interestingly, the transmission rate β is not only considered as the function of time, $\beta_{ij}(t) = \frac{R(t_0)}{\rho(M(t_0))} M_{ij}(t)$, but also modulated by the contact matrix as

Table 3 The use and selection of mobility data for age-structured SEIR-type models

Authors	The use of Google mobility data	Selection of mobility data
[5]	Contact matrix	“Parks”, “Grocery and Pharmacy”, “Workplaces”, “Retail and recreation”, “Transit stations”
[13]	Contact matrix	“Residential”, “Retail and recreation”, “Workplaces”
[19]	Contact matrix	“Retail and recreation”, “Workplaces”
[26]	Contact matrix	“Workplaces”
[35]	Contact matrix	“Retail and recreation”, “Workplaces”, “Grocery and Pharmacy”, “Transit stations”
[36]	The reproduction number	“Retail and recreation”, “Workplaces”, “Grocery and Pharmacy”, “Transit stations”

follows:

$$M_{ij}(t) = \begin{bmatrix} \delta_1 C_{11}(t) & \dots & \delta_1 C_{1n}(t) \\ \vdots & \ddots & \vdots \\ \delta_n C_{n1}(t) & \dots & \delta_n C_{nm}(t) \end{bmatrix},$$

where C denotes an age-group contact matrix, δ is a vector containing susceptibility values relative to each age group. $R(t_0)$ is the reproductive number at time $t = t_0$, and $\rho(M)$ is the spectral radius of a matrix M . More specifically, contact matrix $C(t)$ is modeled as follows:

$$C_{ij}(t) = \omega_h a_h(t) F_{ij}^h + \omega_w a_w(t) F_{ij}^w + \omega_s a_s(t) F_{ij}^s + \omega_c a_c(t) F_{ij}^c,$$

where F^h, F^w, F^s, F^c are household, work, school, and community contact frequency matrices, respectively, derived from the existing literature [25]. The household, workplaces, and community coefficients $a_h(t), a_w(t),$ and $a_c(t)$ are based on percent change in each type of location from its baseline in Google’s COVID-19 community mobility report. Only the school coefficient $a_s(t)$ is set according to the assessed proportion of school openings at each period. The coefficients $\omega_h, \omega_w, \omega_s, \omega_c$ express the number of contacts occurring in different locations above, representing the contribution of each location to the overall contact matrix $C(t)$. The household setting has coefficient of $\omega_h = 1$ since other coefficients are set as relative to ω_h . By data fitting, authors obtain $\omega_w = 1.7, \omega_s = 1.6, \omega_c = 2.9$. Hence, community contact makes the largest contribution to the overall contact matrix $C(t)$. One notable aspect is the asymmetry of their community contact matrix, casting some doubt on the validity of their results.

The common assumption of the normal compartment models is that the population is homogeneous, and it is justified as long as infection within a single community is concerned [16, 20]. It might not work well when a larger scale is concerned [22]. Metapopulation models are created based on a network of subpopulations (i.e., cities, regions, countries) connected by mobility. The disease dynamics inside each patch (i.e., sub-population) follow a compartmental model like those described in the previous section. Metapopulation models always represent socio-technical systems as networks in which nodes describe subpopulations and link the mobility flows between them. More specifically, in the metapopulation model, the mobility data are mainly used to characterize the flow between each metapopulation groups.

Rader et al. [27] used a metapopulation SIR model to study the link between the shape of the epidemic curve and the spatial features of cities. The authors determine the percentage of daily movements within prefectures in China by extracting data on human mobility from the Baidu web platform. The authors extend their results to cities across the world by employing the fitted model from China along with globally extensive covariates. Human mobility data from Baidu are not available for locations outside of China, and hence the authors use the Google mobility dataset to calculate both mobility within shapefile in 310 cities and mobility coming into each city. The authors also mention the limitations of Google’s mobility data that cannot describe population-level mobility patterns. In another paper [28], the authors aimed to determine the extent to which well-planned restrictions relaxing strategies could postpone the resurgence of COVID-19 on a continental scale and curtail community transmission. They first estimate the baseline mobility probability by incorporating mobility data obtained from the pre-COVID-19 continental Google

Table 4 Overview of investigated durations of COVID-19 in studies utilizing Google mobility data in the model

Time length	How many waves	Country	Reference
March 2020-June 2021	3 waves	Kenya	[4]
Feb-May 2020	1 wave	US	[34]
Used 3.19-4.30,2020 to predict 5.1-6.20,2020	1 wave	US	[8]
Used 2.1-9.21, 2020 to predict 9.22,2020-2.28,2021	1 wave	US	[18]
2020.3-2021.1	2 waves	UK, South Africa, Brazil	
3.1.2020-2.23.2021	1 wave	Philippine	[5]
3.12.2020-11.12.2020	1 wave	Ontario, Canada	[19]
May-June 2020	1 wave	France	[26]
3.1.2020-6.1.2020	1 wave	England and Wales	[35]
7.1.2021-11.1.2021	1 wave	Israel	[13]
1.15.2020-2.24.2020	1 wave	China	[27]
1.2020-3.2020 as baseline to simulate 4.4.2020-10.4.2020	Multiple waves but they were simulated under different NPI scenarios	European cities	[28]

NUTS3 (Nomenclature of Territorial Units for Statistics) dataset and Call Data Records from Vodafone in Spain and Italy and then extrapolate it across Europe, employing a linear model, to generate the continental baseline mobility probability. Then the Google COVID-19 data was aggregated to represent the reduction in the NUTS3 area during NPIs. A metapopulation model at the NUTS3 resolution was built. The authors emphasize the importance of incorporating multiple datasets to better capture population-level mobility patterns.

The aforementioned studies did incorporate Google mobility data into their models. Nevertheless, as noted by Unwin et al. [34], relying solely on Google mobility data is insufficient to account for all variations. Although mobility data accounts for a significant portion of the R_t trend, it does not comprehensively depict the evolution of transmission dynamics over time. Other behavioral shifts during COVID-19 are likely contribute to variations as well. Unwin employed a second-order, weekly, autoregressive process to grasp these changes, yet attributing them solely to other transmission determinants or interventions remains challenging. Furthermore, the majority of the aforementioned studies only focused on a single wave, making it unclear how useful they would be for longer time series (refer to Table 4). Currently, there is not sufficient information available to formulate a unified model using Google mobility data for fitting multiple waves in epidemic modeling. Furthermore, in Sect. 3.2, it is worth noting that some authors did not directly incorporate Google mobility data into their mathematical modeling process; instead, they employed it as a validation tool. This highlights the diverse approaches taken in utilizing mobility data across studies.

3.2 Models that deployed the Google mobility data as a tool for validation

Unlike the effort above, several articles did not incorporate Google mobility data into the model. Instead, the authors used it as a tool to check model performance or to validate the model input data. For example, Wong et al. [38] presented a modification of the SIR scheme, considering the long and variable delay times reported in the literature. Forward predictions of the model not only provide robust short-term epidemic estimates (peak position and severity) under social distancing but also the epidemic dynamics later under

releasing orders in the summer of 2020 in Illinois. The effective reproduction number R_t is expressed by the authors as a parametrization involving the basic reproduction number R_0 , a seasonal forcing estimate $F(t)$, a mitigation profile $M(t)$ parametrized as a piecewise cubic Hermite interpolating polynomial, and the susceptible population fraction $S(t)/N$, i.e., $R_t = R_0 F(t) M(t) \frac{S(t)}{N}$. Based on the assumption that no causal relationship exists between R_t and mobility data, even though the model is not supplied with prior information on nonpharmaceutical interventions, it exhibited a mitigation trend that resembles the mobility data reported by Google and Unacast, showing its flexibility and calibration procedure. Google mobility data was not directly deployed in the model-building, instead, it was used as validation tool to compare with the target model results.

Similarly, authors in another paper [9] built a model with Google mobility data, but just to use it as a comparison with their primary model. Based on the SEIR framework, they developed a county-stratified deterministic model using close contact rate to recapitulate the COVID-19 transmission and predict case counts in Connecticut. The close contact rate was derived from the pairs of devices that are within six feet in Connecticut. This close contact rate later is used to determine the mobility metric $M_{contact}(t)$, to parameterize temporal dynamics of transmission parameter $\beta(t)$, where $\beta(t) = \beta_0 M_{contact}(t) \exp[B(t)]$, and $\exp[B(t)]$ is a function that approximates residual changes in transmission parameter. When the estimated value of $[B(t)]$ under a particular mobility metric approximates zero, this mobility metric explains most of the variation in transmission. To evaluate the usefulness of this close contact rate as an input to the transmission model, the authors also fit the SEIR transmission model with mobility metrics from Apple [2], Descartes Labs [11], Facebook [12], Google [1], Cuebiq [10], and with a no-mobility null model. The model with the described close contact rate fits best, and other mobility metrics exhibit a poorer fit. The authors hereby confirm that mobility metrics primarily measure movement, which might not represent close interpersonal contact.

Watson et al. [36] used an age-stratified SEIR model structure to study the dynamics of the SARS-CoV-2 in Damascus, Syria. The time-varying reproduction number R_t is modulated by

$$R_t = R_0 \cdot f(-M_a \cdot (1 - M(t)) - M_\omega \cdot M_a (M(t) - M(t_m)) - \rho_1 - \rho_2 - \dots - \rho_n),$$

where $f(x) = 2 \exp(x) / (1 + \exp(x))$, to capture the impact of mobility data on transmission. $M(t)$ is the inferred mobility throughout the epidemic. ρ_i reflects the change independent of mobility in transmission. However, Google mobility data is not available in Syria, the authors estimate mobility using a Boosted Regression tree model based on an alternative data source. To validate this tree model inferred mobility data, the author compared it with the Google mobility data of Turkey, Iraq, Jordan, Lebanon, and Israel.

Gozzi et al. [14] introduced a metapopulation model with an age structure, employing a stochastic mechanistic epidemic model that considers mobility, physical contacts, and census data. Within this study, the population was subdivided into N comunas and categorized into K age groups. Within each subpopulation, the author employed an *SLIR* compartment model to simulate the dynamics of the epidemic. The author determined reductions in mobility and interpersonal contacts by leveraging data from mobile devices, utilizing this information as an input for the model. Interestingly, while the primary model

proposed by the author did not initially incorporate Google mobility data, it was eventually integrated into an alternative compartmental model presented in supplementary materials. This integration allowed for a comparative assessment of different models. In the model that incorporated Google mobility data, the author treated the entire metropolitan area as a unified, age-structured population. The contacts matrix within this model encompasses a linear combination of four components, representing interactions occurring at school, in the workplace, at home, and in other locations:

$$C(t) = \tilde{\omega}_h(t) \textit{home} + \tilde{\omega}_s(t) \textit{school} + \tilde{\omega}_w(t) \textit{work} + \tilde{\omega}_o(t) \textit{otherlocations},$$

where $\tilde{\omega}(t)$ is the location-specific, time-varying contacts reduction coefficient, and the Google mobility data was used to characterize contacts variations at home, workplace, and other locations. The model with a simplified structure that incorporated Google mobility data actually exhibited poorer performance when compared to the primary model initially proposed.

Chang et al. [7] introduced a metapopulation SEIR model in which subpopulations are from smaller geographic units of the ten largest metropolitan areas in the USA. The subpopulation in these units can interact when visiting a point of interest (POI), which might be a bar, hotel, gym, etc. The system is modeled as a bipartite network with time-varying edges, in which the two types of nodes are units and POI. The weight of an edge $W^{(t)} = W_{ij}$ between a unit and a POI is estimated from SafeGraph data. The researchers used the high Pearson correlation between the SafeGraph and Google mobility datasets to demonstrate the reliability of the SafeGraph datasets since its mobility changes are consistent with Google under the observed period. While the Google mobility data is not directly incorporated into the network, it serves as a validation tool for assessing the reliability of SafeGraph data through its utilization.

4 Conclusions

Before the onset of COVID-19, research on nonpharmaceutical interventions (NPIs) primarily relied on theoretical frameworks, hampered by the notable limitation of lacking empirical data that could describe behavioral changes. However, with the advent of the COVID-19 pandemic, an unprecedented wealth of high-resolution datasets, capturing various facets of NPIs and human mobility, has been amassed and shared. A substantial majority of models now integrate these datasets as inputs. Consequently, there has been a shift from theoretical approaches in the pre-COVID-19 era to data-driven modeling in the post-COVID-19 landscape. Google mobility data, in particular, contributes distinctive and valuable insights into mobility changes and the implementation of interventions, whether integrated into mathematical models or employed as a validation tool. The articles we reference incorporate mobility into models depicting COVID-19 dynamics, often simplified and expressed through contact matrices, contact rates, effective reproduction functions, and the rescaling of key parameters based on mobility data. However, several noteworthy factors deserve attention in evaluating the utility of Google mobility data. While it does capture a significant portion of the R_t trend, it falls short in fully elucidating the dynamics of transmission over time, leaving room for the influence of other behavioral shifts during the pandemic. Unwin et al.'s attempt [34] to capture these dynamics using a second-order, weekly, autoregressive process underscores the complexity of attributing variations solely

to transmission determinants or interventions. Moreover, the focus of most studies on single waves raises questions about the applicability of their findings to longer time series, indicating a need for more robust modeling approaches. Additionally, due to the aggregate nature of Google datasets, there remains a challenge in accounting for individual-level correlations in mobility patterns. The availability of Google's consumer location history feature is also limited to smartphone users, turned off by default, and subject to differential privacy algorithms designed to safeguard user privacy by obscuring fine details. Additionally, mobility estimates may exhibit biases due to the specific populations included in Google mobility data, potentially leading the model to overestimate the spread and resurgence of COVID-19. Consequently, it becomes imperative to broaden the scope by incorporating multiple datasets to capture population-level patterns beyond the confines of any single service or system.

Acknowledgements

The authors would like to thank the editor and the anonymous reviewers for their constructive comments and suggestions to improve the quality of the paper.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by YD, HL, DH, and YZ. The first draft of the manuscript was written by YD, HL, DH, and YZ. Throughout multiple iterations of manuscript revision, each author contributed invaluable insights and suggestions. All authors thoroughly reviewed and unanimously approved the final version for publication.

Funding source

This study was funded by Hong Kong Research Grants Council Collaborative Research Fund (Grant number CRF C5079-21G).

Availability of data and material

Data and material will be made available on request.

Declarations

Declaration of competing interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Received: 21 May 2024 Accepted: 27 June 2024 Published online: 05 July 2024

References

1. Aktay, A., Bavadekar, S., Cossoul, G., Davis, J., Desfontaines, D., Fabrikant, A., Gabrilovich, E., Gadepalli, K., Gipson, B., Guevara, M.: Google COVID-19 community mobility reports: anonymization process description (version 1.1). arXiv preprint [arXiv:2004.04145](https://arxiv.org/abs/2004.04145) (2020)
2. Apple mobility trends reports (2023). <https://covid19.apple.com/mobility>
3. Béraud, G., Kazmierczak, S., Beutels, P., Levy-Bruhl, D., Lenne, X., Mielcarek, N., Yazdanpanah, Y., Boëlle, P.-Y., Hens, N., Dervaux, B.: The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. *PLoS ONE* **10**(7), e0133203 (2015)
4. Brand, S.P., Ojal, J., Aziza, R., Were, V., Okiro, E.A., Kombe, I.K., Mburu, C., Ogero, M., Agweyu, A., Warimwe, G.M.: COVID-19 transmission dynamics underlying epidemic waves in Kenya. *Science* **374**(6570), 989–994 (2021)
5. Caldwell, J.M., de Lara-Tuprio, E., Teng, T.R., Estuar, M., Sarmiento, R.F.R., Abayawardana, M., Leong, R.N.F., Gray, R.T., Wood, J.G., Le, L.V., McBryde, E.S., Ragonnet, R., Trauer, J.M.: Understanding COVID-19 dynamics and the effects of interventions in the Philippines: a mathematical modelling study. *Lancet Reg. Health West. Pac.* **14**, 100211 (2021). <https://doi.org/10.1016/j.lanwpc.2021.100211>
6. Cetron, M., Simone, P.: Battling 21st-century scourges with a 14th-century toolbox. *Emerg. Infect. Dis.* **10**(11), 2053 (2004)
7. Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., Leskovec, J.: Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**(7840), 82–87 (2021)
8. Chiu, W.A., Fischer, R., Ndeffo-Mbah, M.L.: State-level needs for social distancing and contact tracing to contain COVID-19 in the United States. *Nat. Hum. Behav.* **4**(10), 1080–1090 (2020)
9. Crawford, F.W., Jones, S.A., Cartter, M., Dean, S.G., Warren, J.L., Li, Z.R., Barbieri, J., Campbell, J., Kenney, P., Valteau, T.: Impact of close interpersonal contact on COVID-19 incidence: evidence from 1 year of mobile device data. *Sci. Adv.* **8**(1), eabi5499 (2022)
10. Cuebiq mobility insights (2023). <https://www.cuebiq.com/visitation-insights-covid19>
11. Descartes laboratories (2023). <https://www.descarteslabs.com/mobility>

12. Facebook data for good (2023). <https://dataforgood.fb.com/docs/covid19>
13. Gavish, N., Yaari, R., Huppert, A., Katriel, G.: Population-level implications of the Israeli booster campaign to curtail COVID-19 resurgence. *Sci. Transl. Med.* **14**, eabn9836 (2022)
14. Gozzi, N., Tizzoni, M., Chinazzi, M., Ferres, L., Vespignani, A., Perra, N.: Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nat. Commun.* **12**(1), 2429 (2021)
15. Hellewell, J., Abbott, S., Gimma, A., Bosse, N.I., Jarvis, C.I., Russell, T.W., Munday, J.D., Kucharski, A.J., Edmunds, W.J., Sun, F.: Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**(4), e488–e496 (2020)
16. Hethcote, H.W.: The mathematics of infectious diseases. *SIAM Rev.* **42**(4), 599–653 (2000). <https://doi.org/10.1137/S0036144500371907>
17. Hussain, T., Jawed, N., Mughal, S., Shafique, K.: Public perception of isolation, quarantine, social distancing and community containment during COVID-19 pandemic. *BMC Public Health* **22**(1), 1–9 (2022)
18. IHME Team: Modeling COVID-19 scenarios for the United States. *Nat. Med.* **27**(1), 94–105 (2021)
19. Jentsch, P.C., Anand, M., Bauch, C.T.: Prioritising COVID-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study. *Lancet Infect. Dis.* **21**(8), 1097–1106 (2021)
20. Kermack, W.O., McKendrick, A.G.: Contributions to the mathematical theory of epidemics 1. *Bull. Math. Biol.* **53**(1–2), 33–55 (1991). <https://doi.org/10.1007/BF02464423>. Reprinted from 1927
21. Koo, J.R., Cook, A.R., Park, M., Sun, Y., Sun, H., Lim, J.T., Tam, C., Dickens, B.L.: Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect. Dis.* **20**(6), 678–688 (2020)
22. Lipshtat, A., Alimi, R., Ben-Horin, Y.: Commuting in metapopulation epidemic modeling. *Sci. Rep.* **11**(1), 15198 (2021). <https://doi.org/10.1038/s41598-021-94672-w>
23. Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., Edmunds, W.J.: Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**(3), e74 (2008). <https://doi.org/10.1371/journal.pmed.0050074>
24. Prem, K., Cook, A.R., Jit, M.: Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput. Biol.* **13**(9), e1005697 (2017)
25. Prem, K., van Zandvoort, K., Klepac, P., Eggo, R.M., Davies, N.G., Cook, A.R., Jit, M., Dis, C.M.M.I.: Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the COVID-19 era. *PLoS Comput. Biol.* **17**(7), e1009098 (2021). <https://doi.org/10.1371/journal.pcbi.1009098>
26. Pullano, G., Di Domenico, L., Sabbatini, C.E., Valdano, E., Turbelin, C., Debin, M., Guerrisi, C., Kengne-Kueteche, C., Souty, C., Hanslik, T.: Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature* **590**(7844), 134–139 (2021)
27. Rader, B., Scarpino, S.V., Nande, A., Hill, A.L., Adlam, B., Reiner, R.C., Pigott, D.M., Gutierrez, B., Zarebski, A.E., Shrestha, M.: Crowding and the shape of COVID-19 epidemics. *Nat. Med.* **26**(12), 1829–1834 (2020)
28. Ruktanonchai, N.W., Floyd, J., Lai, S., Ruktanonchai, C.W., Sadilek, A., Rente-Lourenco, P., Ben, X., Carioli, A., Gwinn, J., Steele, J.: Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science* **369**(6510), 1465–1470 (2020)
29. SafeGraph (2023). <https://www.safegraph.com>
30. Shi, Y., Wang, Y., Shao, C., Huang, J., Gan, J., Huang, X., Bucci, E., Piacentini, M., Ippolito, G., Melino, G.: COVID-19 infection: the perspectives on immune responses. *Cell Death Differ.* **27**(5), 1451–1454 (2020). <https://doi.org/10.1038/s41418-020-0530-3>
31. The state of the restaurant industry (2023). <https://www.opentable.com/state-of-industry>
32. Tian, S., Hu, N., Lou, J., Chen, K., Kang, X., Xiang, Z., Chen, H., Wang, D., Liu, N., Liu, D., Chen, G., Zhang, Y., Li, D., Li, J., Lian, H., Niu, S., Zhang, L., Zhang, J.: Characteristics of COVID-19 infection in Beijing. *J. Infect.* **80**(4), 401–406 (2020). <https://doi.org/10.1016/j.jinf.2020.02.018>
33. Unacast data for good (2023). <https://www.unacast.com/data-for-good>
34. Unwin, H.J.T., Mishra, S., Bradley, V.C., Gandy, A., Mellan, T.A., Coupland, H., Ish-Horowitz, J., Vollmer, M.A., Whittaker, C., Filippi, S.L.: State-level tracking of COVID-19 in the United States. *Nat. Commun.* **11**(1), 1–9 (2020)
35. Waterlow, N.R., Van Leeuwen, E., Davies, N.G., Flasche, S., Eggo, R.M., CMMID COVID-19 Working Group: How immunity from and interaction with seasonal coronaviruses can shape SARS-CoV-2 epidemiology. *Proc. Natl. Acad. Sci.* **118**(49), e2108395118 (2021)
36. Watson, O.J., Alhaffar, M., Mehchy, Z., Whittaker, C., Akil, Z., Brazeau, N.F., Cuomo-Dannenburg, G., Hamlet, A., Thompson, H.A., Baguelin, M.: Leveraging community mortality indicators to infer COVID-19 mortality and transmission dynamics in Damascus, Syria. *Nat. Commun.* **12**(1), 2394 (2021)
37. Wesolowski, A., Qureshi, T., Boni, M.F., Sundsøy, P.R., Johansson, M.A., Rasheed, S.B., Engø-Monsen, K., Buckee, C.O.: Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proc. Natl. Acad. Sci.* **112**(38), 11887–11892 (2015)
38. Wong, G.N., Weiner, Z.J., Tkachenko, A.V., Elbanna, A., Maslov, S., Goldenfeld, N.: Modeling COVID-19 dynamics in Illinois under nonpharmaceutical interventions. *Phys. Rev. X* **10**(4), 041033 (2020)
39. Yang, W., Shaman, J.: Development of a model-inference system for estimating epidemiological characteristics of SARS-CoV-2 variants of concern. *Nat. Commun.* **12**(1), 1–9 (2021)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.