CHARACTERIZING THE IMPACT OF INTEGRATING SPATIALLY AGGREGATED HUMAN MOBILITY DATA INTO INFECTIOUS DISEASE MODELS

by
Ronan Colbert Corgel

A thesis submitted to Johns Hopkins University in conformity with the requirements for

the degree of Master of Science

Baltimore, Maryland
May 2023

Abstract

When modeling infectious disease dynamics in humans, the spatial scale of mobility data incorporated with models should align with the scale of movement relevant for transmission. While human mobility is a common input of infectious disease models, the effect of spatially aggregating the mobility data on model inferences is rarely explored. In this work, the sensitivity of infectious disease modeling results to different spatial scales of human mobility is considered by integrating three levels of mobility data into metapopulation models, using Sri Lanka as a case study. We observe using mobility data at different levels of spatial aggregation yielded wide variations in estimates of epidemic magnitude and spatial invasion timing. However, differences depended on the introduction location urbanicity, disease transmissibility, and host duration of infectiousness. Despite these findings, when contextualized with COVID-19, the spatial scale of mobility data had little impact on the accuracy of infection arrival times throughout the country. This research highlights the need to investigate multiple mobility data spatial levels when modeling infectious diseases for a better understanding of relevant pathogen transmission scales in addition to potential qualitative differences in conclusions. Our results carry implications for infectious disease modeling best-practices, such that policy decisions made from model inferences are informed by the relevant spatial level of mobility data.

Primary Reader and Advisor: Amy Wesolowski, PhD

Secondary Reader: Kate Grabowski, PhD, ScM

Acknowledgments

I would like to thank my advisor, Dr. Amy Wesolowski, for supporting this work, providing thoughtful guidance throughout my graduate career, and giving me the opportunity to grow as a researcher. I am also appreciative of Dr. Kate Grabowski for serving as my second reader, Dr. Shengjie Lai for sharing access to the Facebook mobility data, the research team in Sri Lanka for all their work with the dengue travel survey, and to colleagues in the Johns Hopkins Infectious Disease Dynamics group for introducing me to the field of infectious disease epidemiology.

Finally, to my friends and family, thank you for celebrating with me for the highs and always being there during the lows. I could not be more grateful for your love and support.

Table of Contents

List of Tables

List of Figures

Introduction

As humans travel throughout the globe, there is potential for infectious diseases to spread in tandem. The COVID-19 pandemic emphasized this phenomenon, as human mobility networks facilitated the speed and breadth at which the virus dispersed across geographies.[1,2] In addition to COVID-19, evidence suggests that disease dynamics for pathogens such as influenza[3,4], malaria[5], dengue[6], yellow fever[7], cholera[8], and measles[9] are influenced by the movement of people from one location to another. Whether pathogens are seeded into naïve regions or contribute to endemic transmission from consistent introductions, human mobility is essential to understand the spatial distribution of infectious disease.

Information on human mobility has conventionally been collected in the form of travel surveys or census migration questionnaires, however these data are typically limited in terms of geographic coverage, time period examined, or populations represented.[10] Technological advancements in the form of mobile phones, GPS loggers, and social media applications allow for human mobility to be captured with high resolution and wide coverage, even in previously data-sparse regions.[11,12] Regardless of the source, all mobility data are spatially resolved and often contain multiple geographic levels of information. Travel survey data can consist of individual details on specific trips accrued or spatially aggregated information of multiple participants. Similarly, mobility data from cell phones and social media applications are generally resolved to administrative levels or tiles due to technology limitations and privacy restrictions.[13,14]

As the spatial granularity of mobility data are altered, interactions between geographies have the potential to change. Often, it is assumed that travel at larger spatial scales is evenly distributed amongst the smaller scales comprising a unit. For instance, suppose two provinces are identified from mobility data as having high inter-provincial travel relative to other areas. More detailed mobility data may reveal that these trips are not evenly distributed within provinces, with inter-provincial travel primarily occurring between the two more populous metropolitan areas of each province. This has significant implications when considering infectious disease spread. Since fine scale patterns of importation, exportation, and infection risk would be masked by aggregated units, specifying a mobility data spatial scale that inadequately captures disease dynamics has consequences for effective outbreak response and informed disease elimination strategies.

While individual movement patterns have been characterized as hierarchical behaviors indicative of travel between neighborhoods, cities, and regions[15], the ideal level of mobility data resolution used to infer disease spread may depend on factors such as pathogen epidemiology, transportation infrastructure, or administrative boundaries. A core component surrounding mobility data spatial scale and infectious disease involves finding a level of detail that sufficiently captures human behavior relevant to disease transmission. Excessively aggregated data potentially ignore essential heterogeneities of travel within units, yet finer mobility scales could fail to maintain average group patterns that reflect individual behavior and therefore lack parsimony.[13,16] For example, coarser mobility data may not be useful in explaining the spatial diffusion of a hyper transmissible pathogen spreading at the local level. Conversely, fine-scale travel information might capture intermediate destinations along a route but lack the ability to identify locations with the highest outbreak potential. Additionally, when multiple spatial levels of mobility data are available, data need to be harmonized with other information such as population estimates or reported cases.[17] To investigate questions involving mobility data spatial resolution and infectious disease, these data must be integrated with infectious disease models.

Infectious disease models have been used during public health emergencies including the COVID-19[18] and Mpox[19] epidemics as well as previous outbreaks of Ebola[20] and SARS.[21] These models proved to be powerful public health tools for disease forecasting, scenario-planning, and decision-making. Data on human mobility or estimates from observed data are typically incorporated in infectious disease models to approximate the spatial spread of disease.[3,6,7,9,22] Despite mobility data commonly serving as inputs of disease transmission models, little is known about how changes in spatial resolution can impact model results. Given that model results are used to inform public health policy, it is critical to determine how sensitive the accuracy and predictive capacity of these results are to mobility data spatial resolution.

Limited availability of high-resolution human mobility data has made questions involving spatial scale challenging to investigate, although previous work has explored similar themes. While past research has examined infectious disease model results using multiple spatial levels of commuting data, spatial levels of mobility data were not explicitly compared and analyses were performed on data from high-income countries.[23] Mobility data have been utilized to estimate the spread of disease in Low or Middle-Income countries (LMICs), nevertheless mobility data were unsuccessful in explaining disease spatial distribution and multiple scales were not considered[24], with the exception of recent work in Ghana.[25] There, the level of detail present in mobility data impacted modeled epidemic spatial spread and peak size, however, the magnitude of change depended on introduction location and disease characteristics.

This research aims to investigate the importance of scale when incorporating human mobility data with infectious disease models, using Sri Lanka as a case study. To evaluate this, three different sources of mobility data were analyzed (mobile phone data, Facebook application mobility data, and travel survey data) and three levels of spatial aggregation were explored (province, district, and divisional secretariat). Mobility data were integrated into SIR metapopulation models and epidemic statistics were compared across spatial resolutions. Differences in epidemic magnitude and spatial invasion were highly reliant on introduction location as well as disease transmissibility and host duration of infectiousness. Finally, infectious disease models utilizing the different mobility data spatial resolutions were compared to COVID-19 incidence data in Sri Lanka, where mobile phone data showed slight superiority in predicting COVID-19 introductions and results did not differ significantly by spatial scale.

Materials & Methods

Population and Geographic Data

Geographic shapefile data for provinces (administrative level 1), districts (administrative level 2), and divisional secretariats (administrative level 3) in Sri Lanka were obtained from the Humanitarian Data Exchange.[26] Shapefiles contained polygon coordinates for 9 provinces, 25 districts, and 339 divisional secretariats. The number of divisional secretariats was inconsistent with the 331 designated by the Sri Lankan government[27] due to the disaggregation of single units into two or three in the shapefile data. Additional information on administrative unit disagreement between datasets is provided in appendix **Table 1**. Divisional secretariat shapefile data was aligned to the Sri Lankan government distinctions and later to divisional secretariats in the mobile phone data.

Data on population counts were downloaded from WorldPop.[28] 2020 constrained UN adjusted data at the 100m resolution level was used to estimate population totals for administrative units at every spatial level. Each tile was adjusted by a common factor[29] so totals reflected 2021 population estimates:

$$adjusted\ population\ tile = population\ tile\ * \frac{Sri\ Lanka\ UN\ Population\ July\ 2021}{Sri\ Lanka\ UN\ Population\ July\ 2020}$$

After adjusting the data, all tiles were assigned to a province, district, and divisional secretariat by performing a spatial join in ArcGIS.[30] Tile centroids that fell within the bounds or were closest to an administrative unit were assigned to the unit.

Mobile Phone Data

Mobile phone data was provided in the form of aggregated and anonymized call data records (CDRs) from a leading telecommunications operator in Sri Lanka, Dialog Axiata. Based on recent financial documentation, the company had over 17.7 million mobile subscribers across all coverage networks in 2021.[31] Call data records are able to document a subscriber's movement when calls or short messaging service (SMS) texts made by the subscriber are routed through the nearest cellular tower. Approximate locations of subscribers can then be estimated by obtaining location information of each cellular tower in the network.[13] Subscriber movement in this dataset was aggregated to the divisional secretariat (administrative level 3) by the data distributor. Mobile phone data captured the number of daily trips between two unique locations, meaning that it was possible for highly mobile subscribers to be included in multiple origin-destination pairs each day.

Daily trips between divisional secretariat origin and destination pairs were available from November 1, 2021 to March 15, 2022. However, the first two weeks of December were missing (December 1, 2021 – December 15, 2021) as well as data from March 6 and 7, 2022. Due to privacy restrictions, routes with 5 or fewer daily trips were censored. In addition to describing movements from unit A to unit B, these data included information on individuals that moved within the administrative level 3 unit. Data was available for nine provinces, 25 districts, and 330 divisional secretariats. Comparable to the geographic data, the number of divisional secretariats did not agree with the Sri Lankan government's distinctions due to the merging of two units, Kothmale East and Kothmale West, in the mobile phone data (appendix **Table 2**). Daily trip counts for each route were adjusted by the data distributor such that trips reflected total population movements rather than the movement of mobile phone subscribers. This process was carried out by scaling the number of trips by the mobile phone operator's market share in the relevant geography.

A data quality issue resulted in 50% decreased trip volume for February and March (appendix **Figures 11a and 11b**). Despite significantly decreased volume, the proportions of trips from common origins to each destination were comparable to previous months (appendix **Figures 11c and 11d**). To correct for this volume inconsistency, trip counts in February and March were scaled upwards so February and March daily route averages matched the corresponding daily route average from January. The formula used for adjustment is described below:

$$adj. daily\ route\ trips = daily\ route\ \ trips * \frac{Jan. daily\ route\ average\ trips}{Feb. or\ Mar. daily\ route\ average\ trips}$$

After applying this formula, February and March volume across routes reflected volume from previous months. Following this process, mobile phone data was aggregated to the province and district levels to create three resolutions of human movement. Then, all spatial levels of data were averaged across the entire time frame to obtain a daily average number of trips for each origin-destination pair.

Facebook Mobility Data

Data on the mobility of Facebook application users in Sri Lanka was obtained from Meta's Facebook Disaster Map project.[32] These data represented movements of Facebook application users that opted into sharing their location history. Mobility information was available from March 30, 2020 to May 21, 2022 for nine provinces, 25 districts, and 324 divisional secretariats across the country. In contrast to the mobile phone data, Facebook mobility data were available at eight-hour intervals and in two different aggregations, Bing Tile[33], the resolution system used by Bing Maps and Microsoft, and province (administrative level 1). User movements between tiles and provinces were calculated by comparing the most common location of the user within an 8-hour period to the most common location of the user within the consecutive 8-hour period.[14] Human movements within units were also recorded in the process. Tile data were aggregated to the administrative level 3 and level 2 units by assigning the centroid of each tile to an administrative unit using ArcGIS.[30] Following the assignment of tiles to units, province, district, and divisional secretariat movement data were aggregated from 8-hour trip counts to daily trip counts and averaged across all days in the time period.

Dengue Travel Survey Data

Individuals admitted to the Clinical Centre for Managing Dengue and Dengue Hemorrhagic Fever (CCMDDHF) at the District General Hospital in Negombo, Sri Lanka were eligible for participation in the travel survey. Participants were enrolled in the study from February 2020 to December 2021 and completed travel surveys covering locations visited in the two weeks prior to dengue infection. Origin-destination pairs without complete location information were excluded, resulting in data on 658 trips from 308 individuals. Geolocation information was obtained by using the Google Geocoding API, converting residential and destination addresses to latitude and longitude coordinates.[34] Coordinate data was then used to assign administrative units for each origin and destination in ArcGIS[30] and individual trips were aggregated to the divisional secretariat, district, and province levels.

COVID-19 Incidence Data

COVID-19 situation reports containing information on cumulative cases by health district were available on the Sri Lankan Government's Epidemiology Unit website.[35] The Epidemiology Unit, under the Ministry of Health, punished daily updates on the COVID-19 situation in Sri Lanka beginning January 28, 2020. Information on each daily update was compiled into time-series data from November 14, 2020 to May 29, 2021. Prior to November 14, cases were sporadically reported with many not assigned to a health district. Due to these reporting errors, the COVID-19 Alpha variant wave was examined instead of earlier fluxes in cases. Case data were only examined until May 29 because the Alpha variant wave began in early April 2021 and reached a peak in late May 2021 (appendix **Figure 13**). Cumulative COVID-19 case data were provided by health district, which corresponded to districts with the addition of Kalmunai, a city in Ampara district that reported cases independently. Cumulative case data was converted to incident case data by subtracting the previous day's case total from the index case total and a 7-day rolling average was calculated to smooth incident cases across time.

Descriptive Analyses

All mobility data (mobile phone, Facebook, and travel survey) were aggregated and averaged across time to create average trip counts for each origin-destination pair at the divisional secretariat, district, and province levels. Trip proportion was calculated as the fraction of trips to a destination out of all trips from a common origin:

$$Trip\ Proportion = \frac{Trips_{ij}}{\sum_{j=1}^{n} Trips_{ij}}$$

For example, if there were 100 total trips out of Colombo and 25 went to Gampaha, then the trip proportion would be 0.25 for that pair. Origin-destination matrices were created from these trip proportions for each dataset and administrative level, resulting in a total of nine origin-destination matrices. Trip proportions along with the proportion of trips occurring within an administrative unit were compared across mobility data sources and spatial levels. Additionally, external trip proportions were compared between nested units at different spatial levels. Data missingness was evaluated using univariate logistic regression, with missingness as a binary outcome and origin population, destination population, and trip distance as explanatory variables.

Modeling Analyses

Disease dynamics at different spatial levels were simulated using a discrete-time stochastic metapopulation SIR model.[36,37] In this mechanistic model, individuals transitioned between susceptible, infected, and recovered compartments, analogous to the chronological transition of humans between disease states. The metapopulation model accounted for spatial disease dynamics by treating each geographic unit as its own patch. The number of patches fluctuated depending on the spatial resolution level of the model. For instance, the district-level metapopulation model contained 25 patches, equivalent to the 25 districts in Sri Lanka. The equations of each compartment in the model are listed below for patch, k, at time t:

$$S_{k,t} = S_{k,t-1} - \iota_{k,t}$$

$$I_{k,t} = I_{k,t-1} + \iota_{k,t} - \rho_{k,t}$$

$$R_{k,t} = R_{k,t-1} + \rho_{k,t}$$

Where S is the number of susceptible individuals, I is the number of infected individuals, R is the number of recovered individuals, ι is the number of new infections, and ρ is the number of new recoveries. Frequency dependent disease transmission, untethered from population size, was assumed to isolate the effect of mobility data spatial aggregation on modeling results.

Travel was incorporated into the model as an origin-destination matrix containing trip proportions for the specified spatial resolution. Stochasticity was added into the travel process by drawing from a Binomial distribution with the number of trials equal to the unit origin population, N_i. The probability of success was assigned as the observed proportion of all trips from origin i that went to destination j, $\Pi_{i,j}$, to estimate the number of trips for the route at each time step, $\pi_{i,j,t}$:

$$\pi_{i,j,t} \sim Binomial(N_i, \Pi_{i,j})$$

This process was completed for the entire observed travel matrix to obtain an estimated number of trips for every origin-destination pair. The matrix was then normalized such that each cell reflected the proportion of trips traveling to a destination from a common origin at each time step, $T_{i,j,t}$.

After incorporating stochasticity into the travel process, travel was factored into the stochastic nature of infection. The probability of infection was allowed to fluctuate with the number of infections in patch k as well as the number of infections in all other patches, proportional to the amount of travel into patch k. A Binomial distribution was used to estimate the number of new infectious occurring in patch k, based on the number of susceptible individuals at the previous time step in patch k and the probability of infection:

$$\iota_{k,t} \sim Binomial(S_{k,t-1}, 1 - e^{-\delta t \beta \left(T'_{i,k,t} \times \frac{I_{k,t-1}}{N_k}\right)})$$

Where δt is the time step amount, β is the transmission coefficient, $T'_{i,k,t}$ represents the transposed normalized travel matrix containing the proportion of trips from each origin to patch k at time step t, and N_k is the patch population. Similarly, stochasticity was added into the recovery process assuming a Binomial distribution, where γ is the rate of recovery:

$$\rho_{k,t} \sim Binomial(I_{k,t-1}, 1 - e^{-\delta t \gamma})$$

Disease dynamics produced by the metapopulation model were examined on a daily time step, for 365 days, following the introduction of a single individual infected with a novel pathogen into Sri Lanka. Based on the model structure and brief time-period explored, recovered individuals were assumed to be immune, ineligible for reinfection. In these modeling analyses, the proportion of the population susceptible was set at 0.90 for all metapopulation patches, reflecting a population with little existing immunity. Births and deaths were not factored into the model, assuming a constant population size over the year-long period. Additionally, the model designated a similar rate of contact between subpopulations and homogeneous mixing within each patch.

Mobile phone data was incorporated into the model to approximate travel between patches. While Facebook mobility data was examined in the descriptive analyses, it was not used for infectious disease modeling due to high levels of missingness across spatial resolutions. Travel survey data was also not incorporated into the model due to the limited geographic scope of the data, as enrollment only occurred at one hospital in the country. Missing trip count information in mobile phone data due to censoring was assumed to be zero. To examine changes in infectious disease modelling results between different mobility data spatial resolutions, simulations of the metapopulation model were carried out at the province, district, and divisional secretariat levels.

A spectrum of introduction locations, R_0 values, and durations of infectiousness were examined to explore their sensitivity to modeling result differences across mobility data spatial resolutions. For each modelling scenario at a given spatial resolution and combination of introduction location, R_0, and infectiousness duration, 50 simulations were run to estimate the average number of incident cases for each unit at all time steps. Epidemic magnitude and spatial invasion statistics were then calculated from the average result across simulations. Epidemic magnitude was calculated as the sum of incident infections that occurred throughout the year-long time-period. Spatial invasion, on the other hand, was estimated as the average time one cumulative case was observed in each unit (excluding the introduction unit). Epidemic magnitude and spatial invasion were the primary statistics compared between model scenarios.

COVID-19 Analyses

Epidemic magnitude and spatial invasion timings were investigated for the COVID-19 Alpha variant wave between observed data and modeled estimates from April 9, 2021 to May 29, 2021.

April 9 was selected as the model start date since Sri Lanka reached a local minimum number of cases that day and the Alpha variant became dominant in April 2021 (appendix **Figure 13**). The metapopulation model described in the previous section was used to simulate COVID-19 dynamics in Sri Lanka at the district and province levels (administrative levels 2 and 1). To assess the usefulness of mobility data in explaining disease dispersion, observed COVID-19 data was compared with model results that incorporated mobility using mobile phone data in addition to model results assuming no mobility. Scenarios with different levels of spatial resolution were also examined. Mobile phone data at the administrative level 3 resolution were not utilized since case reporting occurred at administrative level 2.

The total number of reported cases in Sri Lanka on April 9 was 83,489, representing less than 1% of the population. To account for underreporting of cases, the country was assumed to be 95% susceptible. The R package EpiEstim was used to approximate R_0 during the simulation period, which was indicated as 1.5.[38] The effective reproduction number was estimated for the entire country in addition to specific districts (appendix **Figure 14**). The duration of infectiousness, or $\frac{1}{\gamma}$ (inverse rate of recovery), was estimated at 10 days from literature.[39,40,41] At the start of the modeling period, a cumulative total of the observed COVID-19 cases from the previous nine days in each unit were assigned to the infected compartment to estimate the number of currently infectious individuals.

Since COVID-19 was actively circulating during the study period, spatial invasion was estimated as the time a spatial unit reached pre-specified incident case thresholds. The spatial invasion timing case threshold was set at 5 incident cases a day for district-level scenarios and 25 incident cases a day for province-level scenarios. Epidemic statistics of magnitude and spatial invasion were compared between observed COVID-19 data and modeling results assuming mobility or no mobility. Root mean square error (RMSE) was calculated to determine the accuracy of each model and spearman rank correlations were computed to evaluate spatial invasion ordering.

Code & Data Availability

All analytics were performed in R version 4.2.1[42] and ArcGIS Pro version 2.9.[30] The code used to analyze mobility data, run infectious disease models, and produce figures is located on GitHub.[43] Geographic boundary, population, and COVID-19 incidence data are all publicly available through the indicated sources. While original mobility data are not permitted to be shared, simulated mobile phone and Facebook mobility data are also provided on GitHub.

Results

Descriptive Differences Observed across Mobility Datasets and Administrative Levels

Sri Lanka's nine provinces, 25 districts, and 330 divisional secretariats were visualized in **Figure 1** along with each administrative level's population distribution. The average unit population for administrative level 3 was 55,748, ranging from 673 to 303,400, while administrative levels 2 and 1 had average population sizes of 735,874 (ranging from 40,100 to 2,400,000) and 2,044,095 (ranging from 932,900 to 6,019,000), respectively. Across all administrative levels, population distributions remained right-skewed with highly populated administrative units concentrated around the capital, Colombo. For administrative level 3, only four units had a population of over 250,000 yet more than 85% of units had less than 100,000 residents (**Figure 1c**). Colombo and a rural town in northern Sri Lanka, Madhu, were highlighted in each map of **Figure 1** since modeling analyses focused on these areas.

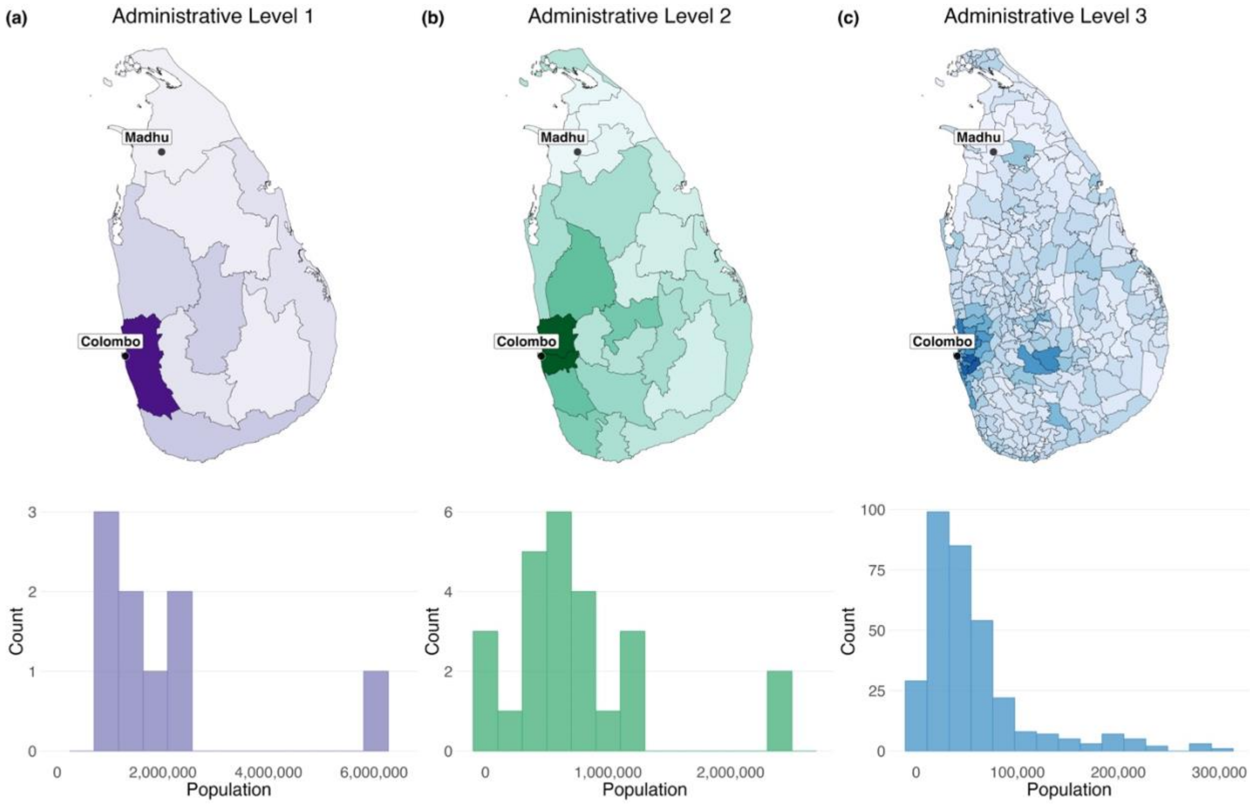

Figure 1. Sri Lanka Geographic Boundaries and Population Distributions by Administrative Level

(a) Map and histogram of 2021 population distribution for provinces (administrative level 1). **(b)** Map and histogram of 2021 population distribution for districts (administrative level 2). **(c)** Map and histogram of 2021 population distribution for divisional secretariats (administrative level 3). In each map, the capital Colombo and the rural town Madhu are highlighted.

Across all administrative units and mobility data sources, there was a negative relationship between the logarithm number of trips and trip distance among an origin-destination pair (**Figure 2**). That is, as distance between two locations increased, the logarithm number of trips for the route decreased. The slope of this relationship was not consistent across datasets or administrative levels, however. For administrative level 3, the slope of the relationship was -0.011 for mobile phone data, -0.115 for Facebook data, and -0.005 for travel survey data. For mobile phone data, the slope of the relationship was -0.028 for administrative level 1, -0.024 for administrative level 2, and -0.011 for administrative level 3 (appendix **Table 3**).

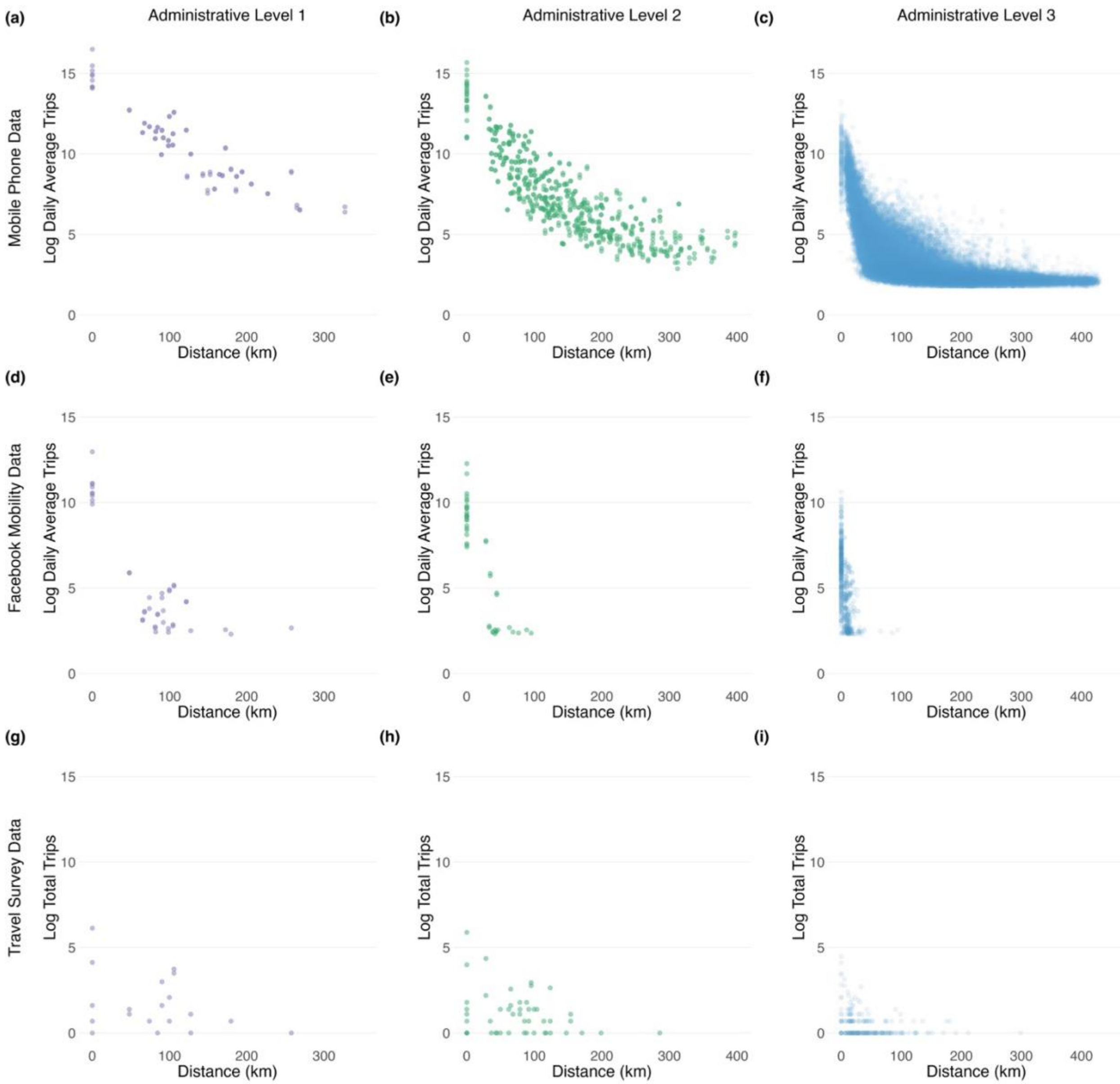

Figure 2. Route Logarithm Trip Counts by Mobility Dataset and Administrative Level

(a), (b), (c) The relationship between logarithm of the average daily trip count and distance for each origin-destination pair in the mobile phone data at administrative levels 1, 2, and 3. **(d), (e), (f)** The relationship between logarithm of the average daily trip count and distance for each origin-destination pair in the Facebook mobility data at administrative levels 1, 2, and 3. **(g), (h), (i)** The relationship between logarithm of the total trip count and distance for each origin-destination pair in the travel survey data at administrative levels 1, 2, and 3.

Slopes varied due to characteristic differences across mobility datasets and administrative levels. For instance, at administrative level 3, mobile phone data recorded the largest number of trips (40,987,000 average daily trips compared to 630,500 for Facebook data and 658 total trips for travel survey data) and travel survey data had the longest trips on average (21.6 km compared to 17.8km in mobile phone data and 0.5 km for Facebook data). For mobile phone data, as mobility data became coarser, the maximum number of daily trips per route increased from 568,700 to over 14,700,000 and the average trip distance decreased from 17.8 km to 9.2 km (appendix **Table 4**).

Origin-destination matrices of trip proportion were compiled for the nine variations of mobility data source and administrative level (**Figure 3**). Trip proportion was highest along the diagonal of each matrix, where cells represented the proportion of trips occurring within the unit, rather than traveling to another unit. For mobile phone data at administrative level 2, the average trip proportion for trips occurring within a district was 0.773 opposed to an average proportion of 0.001 for trips to another district (compared to 0.881 vs. 0.015 for administrative level 1 and 0.227 vs. 0.002 for administrative level 3). In finer resolution matrices, regional travel between neighboring units was depicted by clusters of high trip proportions along the diagonal, similar to the relationships from **Figure 2** where larger trip magnitudes were observed for locations closer together.

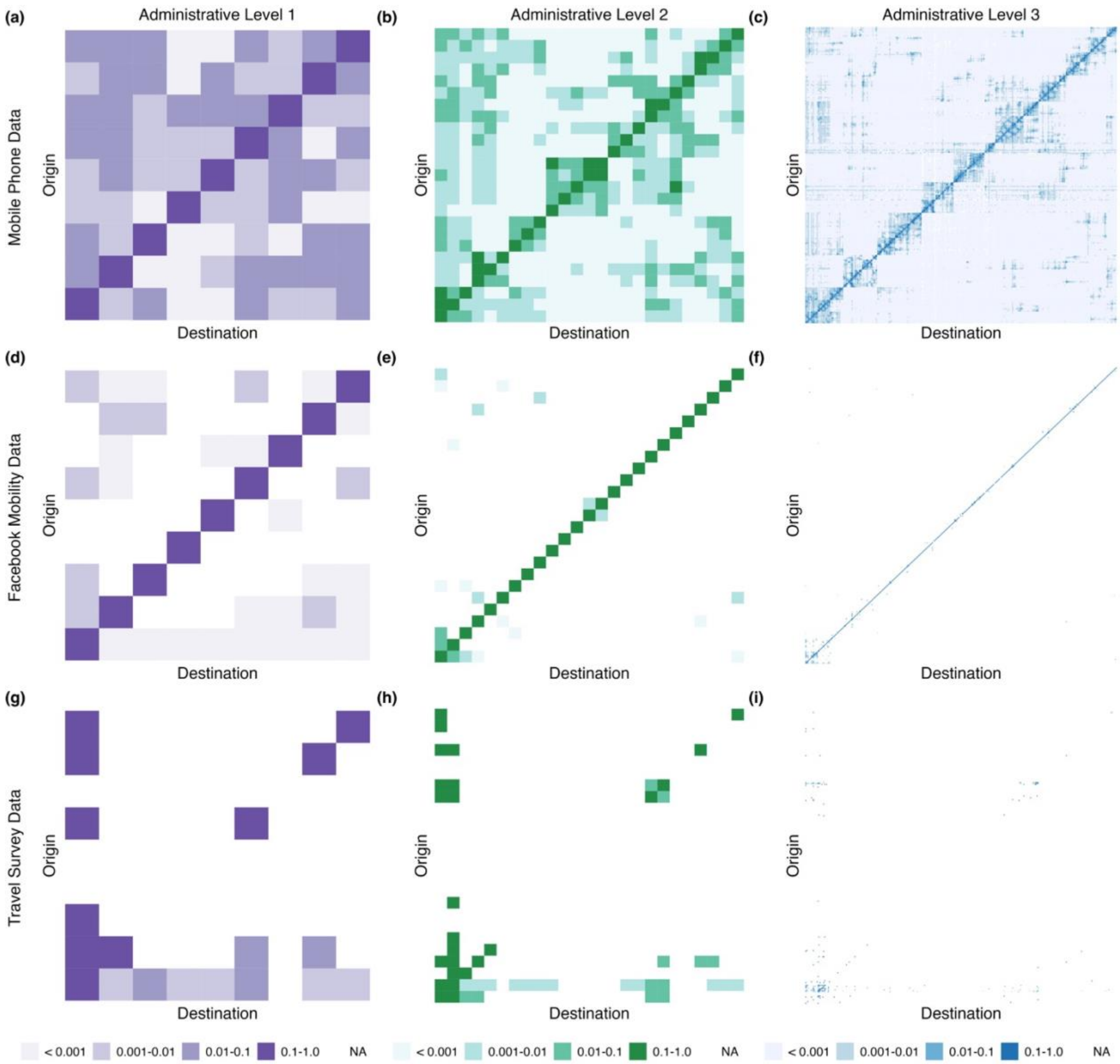

Figure 3. Trip Proportion Origin-Destination Matrices by Mobility Dataset and Administrative Level

(a), (b), (c) Daily average trip proportion for mobile phone data at administrative levels 1, 2, and 3. **(d), (e), (f)** Daily average trip proportion for Facebook mobility data at administrative levels 1, 2, and 3. **(g), (h), (i)** Trip proportion for travel survey data at administrative levels 1, 2, and 3.

Mobile phone data matrices had the smallest amount of missingness in terms of the proportion of origin-destination pairs with observed trips. For administrative level 1 data, 0% of cells were missing for mobile phone data compared to 50.6% for Facebook data and 76.5% for travel survey data, while missingness for administrative level 2 was 0%, 92.6%, and 92.3% and administrative level 3 had missingness of 1.8%, 99.4%, and 99.8% for each mobility source. Univariate logistic regression showed that smaller origin and destination populations as well as longer trip distances were associated with higher odds of missingness across all administrative levels and mobility data sources (appendix **Table 5**). For each administrative level and mobility source with missing data, origin population, destination population, and trip distance were significantly associated with missingness, aside from destination population in Facebook administrative level 1 data.

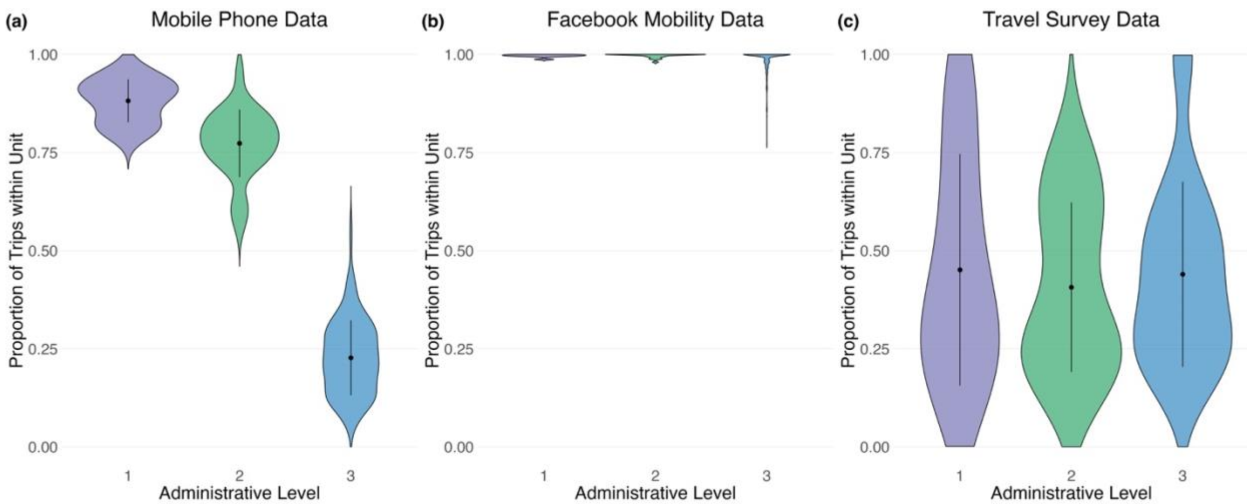

Figure 4. Internal Trip Proportion Distribution by Mobility Dataset and Administrative Level

(a) Internal trip proportion distributions for mobile phone data administrative levels 1, 2, and 3. **(b)** Internal trip proportion distributions for Facebook mobility data administrative levels 1, 2, and 3. **(c)** Internal trip proportion distributions for travel survey data administrative levels 1, 2, and 3. Points represent the mean internal trip proportion for each distribution and lines show one standard deviation above and below the mean.

Internal trip proportions examined the fraction of trips that occurred within the unit out of all trips originating from the unit. In the mobile phone data, as spatial level moved from finest resolution to coarsest, the mean internal trip proportion also increased from 0.227 for divisional secretariates, to 0.773 for districts and 0.882 for provinces (**Figure 4a**). These proportions had a positive correlation with population size, meaning that units with larger populations tended to have a high proportion of people traveling within the unit (appendix **Figure 15**). Heterogeneities observed between levels in the mobile phone data were not present in the Facebook data, where the internal trip proportion was near 1.00 across all spatial aggregation levels (**Figure 4b**). While travel survey data did exhibit similar trends as the mobile phone data in terms of a positive correlation between internal trip proportion and population size, there were wide distributions with no clear aggregation pattern from administrative levels 3 to 1 (**Figure 4c**).

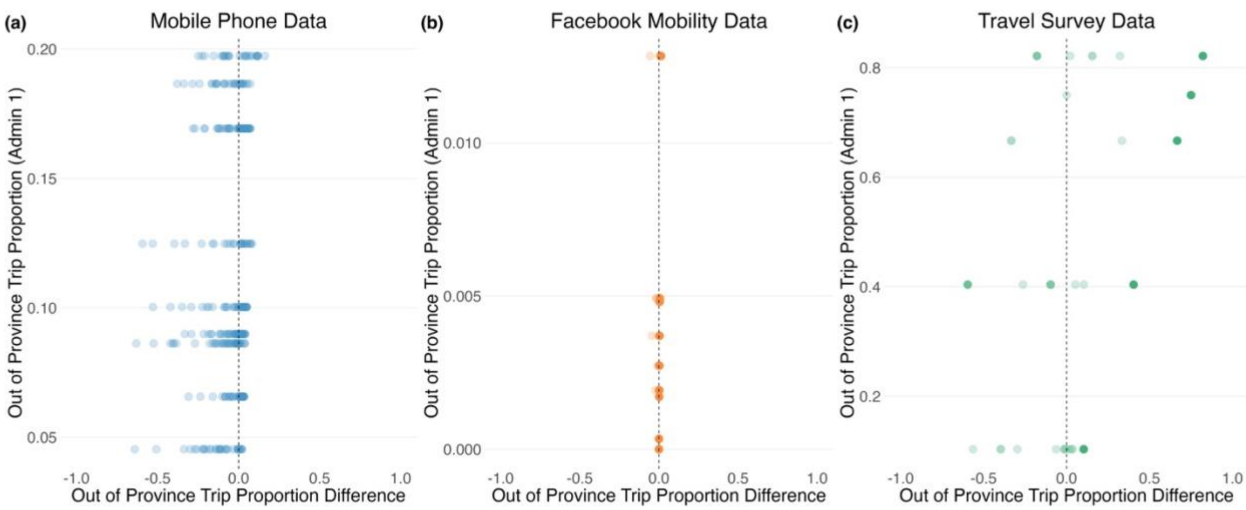

Figure 5. Differences in External Travel between Administrative Levels 1 and 3

(a), (b), and (c) Difference in out of province travel between administrative 1 units and nested administrative level 3 units by administrative level 1 out of province trip proportion for mobile phone, Facebook, and travel survey data. Trip proportion differences were calculated by subtracting administrative level 3 out of province trip proportion from administrative level 1 out of province trip proportion.

To determine potential causes for discrepancies in modeling results between mobility data resolutions, differences in external trip proportions were examined. For administrative levels 1 and 3, the proportion of trips occurring out of the province for an administrative level 1 unit was compared to the proportion of trips occurring out of the province for a nested administrative level 3 unit. The differences between administrative units depended on population, however this was not consistent for each administrative level pairwise grouping. For administrative levels 1 and 3 and administrative levels 2 and 3 (**Figure 6b** and **Figure 6c**), a nested unit with a log population size of eight on average had 25% more external trips than their parent unit, while a nested unit with a log population size of 12 on average had slightly less external travel than their parent unit. These trends did not hold for comparisons between administrative levels 1 and 2, where little difference in external travel was observed between nested and parent units (**Figure 6a**). Due to the high levels of missingness in Facebook and travel survey data, only mobile phone data was examined in **Figure 6** and the following modeling analyses, however, magnitude of external trip proportion differences between nested units varied by mobility data source (**Figure 5** and appendix **Figure 16**).

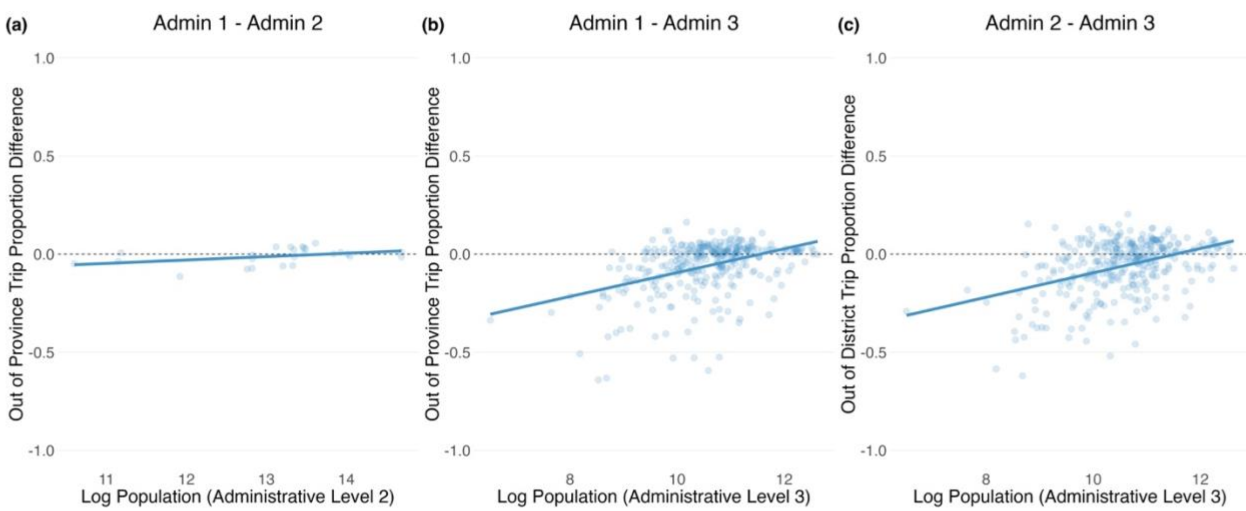

Figure 6. Differences in External Travel between Administrative Levels by Population for Mobile Phone Data

(a) Difference in out of province trip proportion between administrative levels 2 and 1 by logarithm district population. **(b)** Difference in out of province trip proportion between administrative levels 3 and 1 by logarithm divisional secretariat population. **(c)** Difference in out of district trip proportion between administrative levels 3 and 2 by logarithm divisional secretariat population. The trend line in each figure represents a linear fit to the data.

Model Result Differences between Mobility Spatial Levels Varied by Introduction Location, Transmissibility, and Duration of Infectiousness

Disease dynamics were simulated for each level of mobile phone mobility data and epidemic statistics were compared between levels and across introduction locations (**Figure 7**), transmissibility values (**Figure 8**), and durations of infectiousness (**Figure 9**). Introduction locations of smaller population size were more likely to produce larger outbreaks and more rapid spatial invasion timings with finer scale mobility data than corresponding models using coarser mobility data. Examining administrative levels 3 and 1 and a pathogen introduction into the unit with the smallest population size, fine resolution mobility data simulated an epidemic with 5 million more cases and average spatial invasion occurring 34 days sooner than coarser mobility data (**Figure 7b** and **Figure 7e**). These differences decreased and even reversed as introduction location population grew. For a novel pathogen introduction into the largest population unit, fine resolution mobility data simulated an epidemic with 1.3 million fewer cases and average spatial invasion occurring 5 days later than coarser mobility data. Between administrative levels 1 and 2, these trends were less pronounced. For these administrative levels, epidemic magnitude and spatial invasion estimates fluctuated by approximately 2 million cases and 30 days, opposed to 5 million cases and 60 days for model comparisons between different administrative level pairs.

While a clear pattern emerged between model result differences and introduction location population, no such trend was observed when considering differences in modeled epidemic magnitude and transmissibility for an urban introduction location. Comparing administrative levels 1 and 2, fine resolution mobility data produced 1.6 million more cases than coarse mobility data when R_0 was set at 3.9 and 2.2 million less cases when R_0 was 4.0 (**Figure 8a**). On the other hand, the rural introduction location simulated larger epidemics with finer scales of mobility data across 100% of R_0 values explored. For both urban and rural introduction locations, differences in average spatial invasion timing between models simulated with varying mobility data levels decreased as transmissibility increased. Examining administrative levels 1 and 3, the gap between average spatial invasion timing decreased from 32 to 8 days for the rural introduction and from 4 days to 1 day for the urban introduction as R_0 increased from 1.3 to 4.0 (**Figure 8e**).

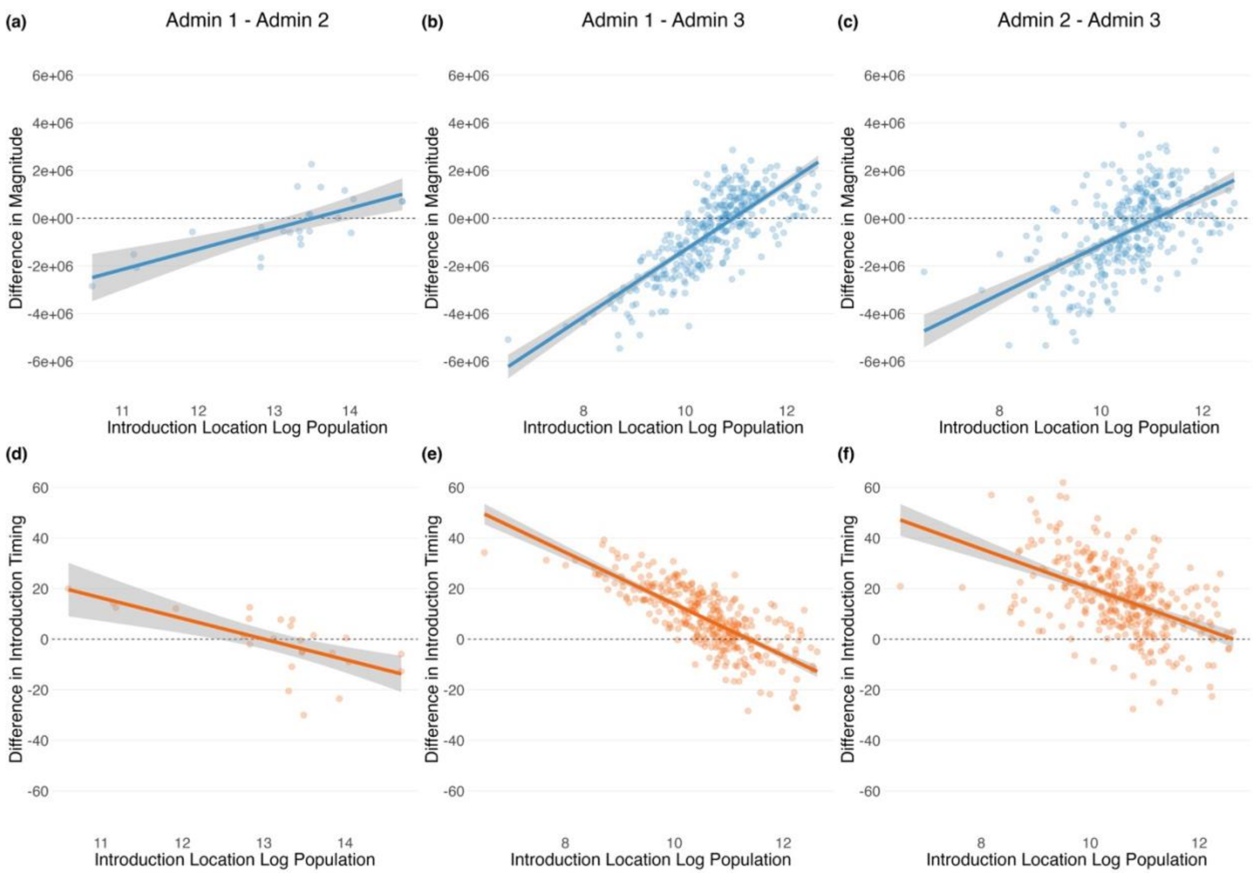

Figure 7. Differences in Epidemic Statistics between Administrative Levels by Introduction Location

(a), (d) Differences in epidemic magnitude and introduction timing between administrative level 2 mobility data model results and administrative level 1 model results, by logarithm population of the introduction location. **(b), (e)** Differences in epidemic magnitude and introduction timing between administrative level 3 mobility data model results and administrative level 1 model results, by logarithm population of the introduction location. **(c), (f)** Differences in epidemic magnitude and introduction timing between administrative level 3 mobility data model results and administrative level 2 model results, by logarithm population of the introduction location. Trend lines represent a linear fit to the data and shaded regions show the 95% confidence interval. Epidemic statistic differences are the difference between two modeling scenarios, with each scenario averaged across 50 simulations. Duration of infectiousness is set at 7 days and R_0 is set at 1.5 for all simulations.

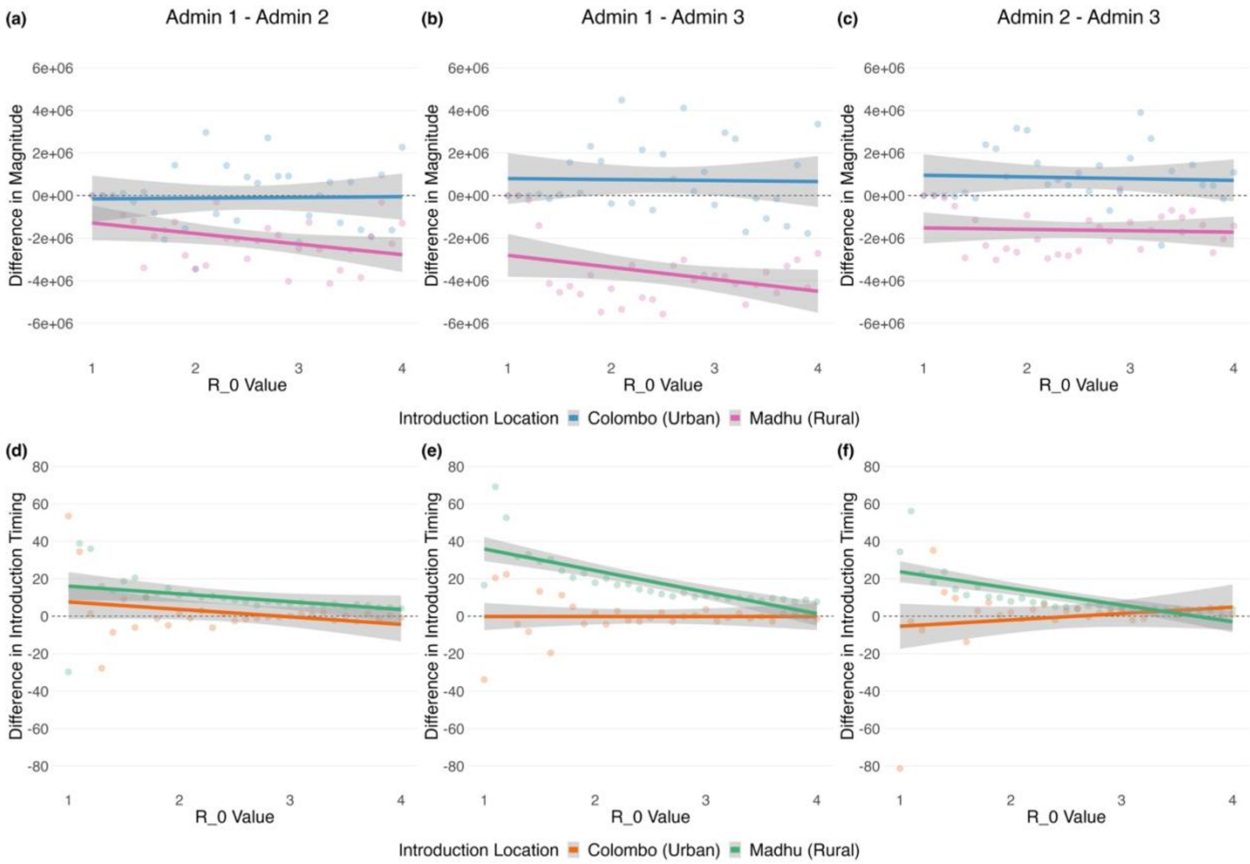

Figure 8. Differences in Epidemic Statistics between Administrative Levels by R_0 for an Urban and Rural Introduction Location

(a), (d) Differences in epidemic magnitude and introduction timing between administrative level 2 mobility data model results and administrative level 1 model results, by R_0 value. **(b), (e)** Differences in epidemic magnitude and introduction timing between administrative level 3 mobility data model results and administrative level 1 model results, by R_0 value. **(c), (f)** Differences in epidemic magnitude and introduction timing between administrative level 3 mobility data model results and administrative level 2 model results, by R_0 value. Trend lines represent a linear fit to the data and shaded regions show the 95% confidence interval. Epidemic statistic differences are the difference between two modeling scenarios, with each scenario averaged across 50 simulations. Duration of infectiousness is set at 7 days for all simulations.

As the duration of infectiousness increased, differences in epidemic magnitude between models using contrasting mobility data resolutions decreased regardless of introduction location type. However, for lower levels of duration and a rural introduction location, finer resolutions of mobility data produced larger epidemics relative to coarser mobility scales while finer mobility resolutions typically resulted in smaller epidemics when an urban introduction was considered. This is illustrated in **Figure 9b**. As duration of infectiousness increased from 1 day to 20 days, epidemic magnitude difference increased from -4 million to -50,000 for the rural introduction and decreased from 3.5 million to 1,200 for the urban introduction location. When examining duration of infectiousness and spatial invasion timing patterns in the rural introduction location, model results from finer mobility data produced more rapid epidemics than coarse mobility data as duration of infection increased. Comparing administrative levels 2 and 3, the difference in average time of spatial invasion increased from 2 to 28 days as duration of infectiousness increased from 1 day to 20 days. For the urban introduction location, there did not appear to be a clear trend for spatial invasion timing and duration of infectiousness.

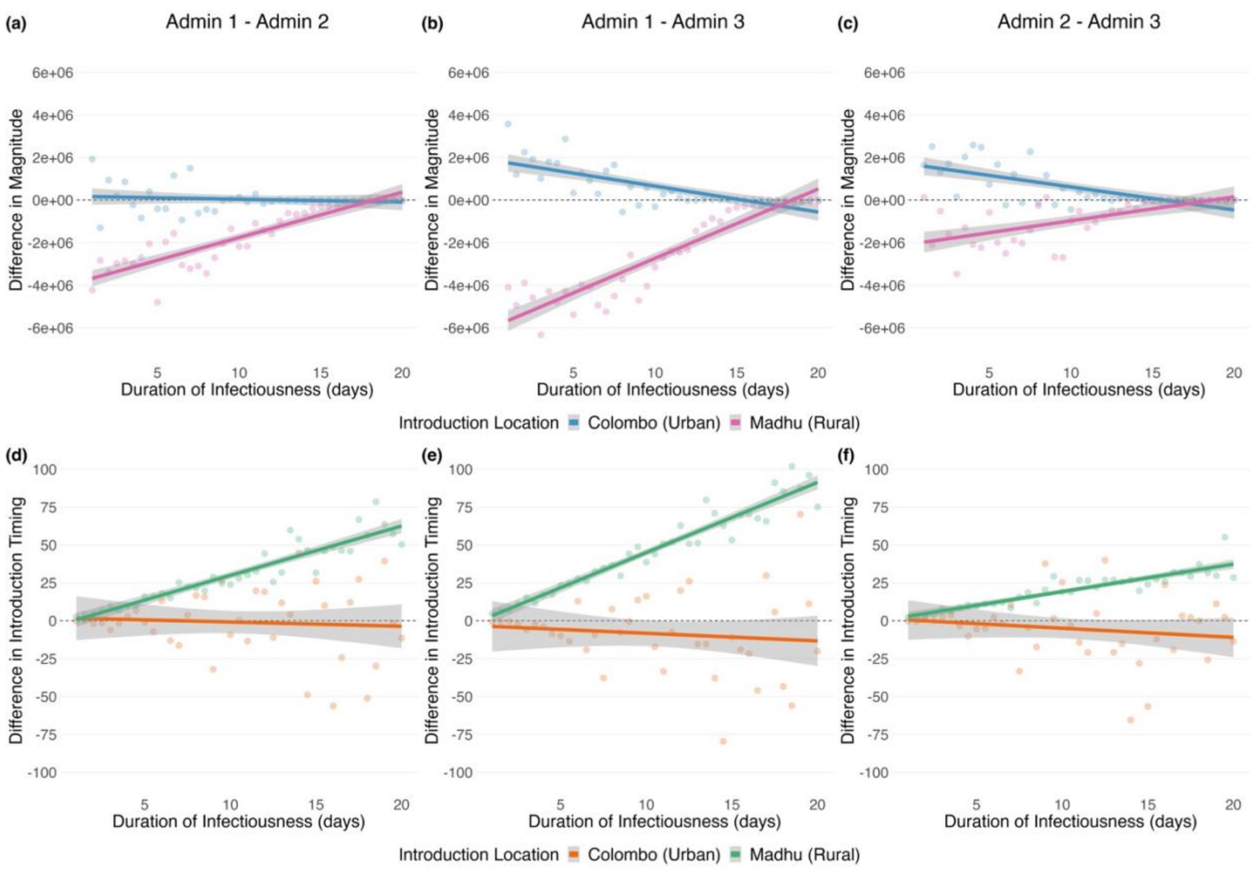

Figure 9. Differences in Epidemic Statistics between Administrative Levels by Duration of Infectiousness for an Urban and Rural Introduction Location

(a), (d) Differences in epidemic magnitude and introduction timing between administrative level 2 mobility data model results and administrative level 1 model results, by duration of infectiousness. **(b), (e)** Differences in epidemic magnitude and introduction timing between administrative level 3 mobility data model results and administrative level 1 model results, by duration of infectiousness. **(c), (f)** Differences in epidemic magnitude and introduction timing between administrative level 3 mobility data model results and administrative level 2 model results, by duration of infectiousness. Trend lines represent a linear fit to the data and shaded regions show the 95% confidence interval. Epidemic statistic differences are the difference between two modeling scenarios, with each scenario averaged across 50 simulations. R_0 is set at 1.5 for all simulations.

Mobility Data Scale had Little Effect on Accuracy of COVID-19 Epidemic Magnitude and Spatial Invasion Timing

Models simulated at administrative levels 1 and 2, with and without incorporating mobility into the model, were compared to observed COVID-19 incidence data from Sri Lanka situation reports. When examining spatial invasion, the distribution of introduction timing differences between modeled estimates and observed data was narrower for models using mobile phone data than models assuming no mobility (**Figure 10a**). Introduction timing differences ranged from -11 to 19 and -16 to 15 days for administrative levels 1 and 2 models that incorporated mobility and from -17 to 22 and -18 to 21 days for administrative levels 1 and 2 models that did not include mobility. Additionally, introduction timing differences did not follow any trend as observed introduction time or unit population size changed (appendix **Figure 17**).

Regarding the error of model results, the administrative level 1 model assuming mobility had the lowest mean RMSE across 1,000 simulations (9.63) followed by administrative level 2 with mobility (9.75), administrative level 2 with no mobility (13.30), and administrative level 1 with no mobility (13.57). While mean RMSE values were lower for models assuming population movement, there was substantial overlap of the intervals covering 95% of RMSE simulation values between movement and no movement models (**Figure 10b**). Furthermore, no administrative level or mobility assumption model appeared to have superiority in predicting spatial invasion ordering, with mean spearman rank correlation coefficients between 0.66 and 0.69 for all models (**Figure 10c**).

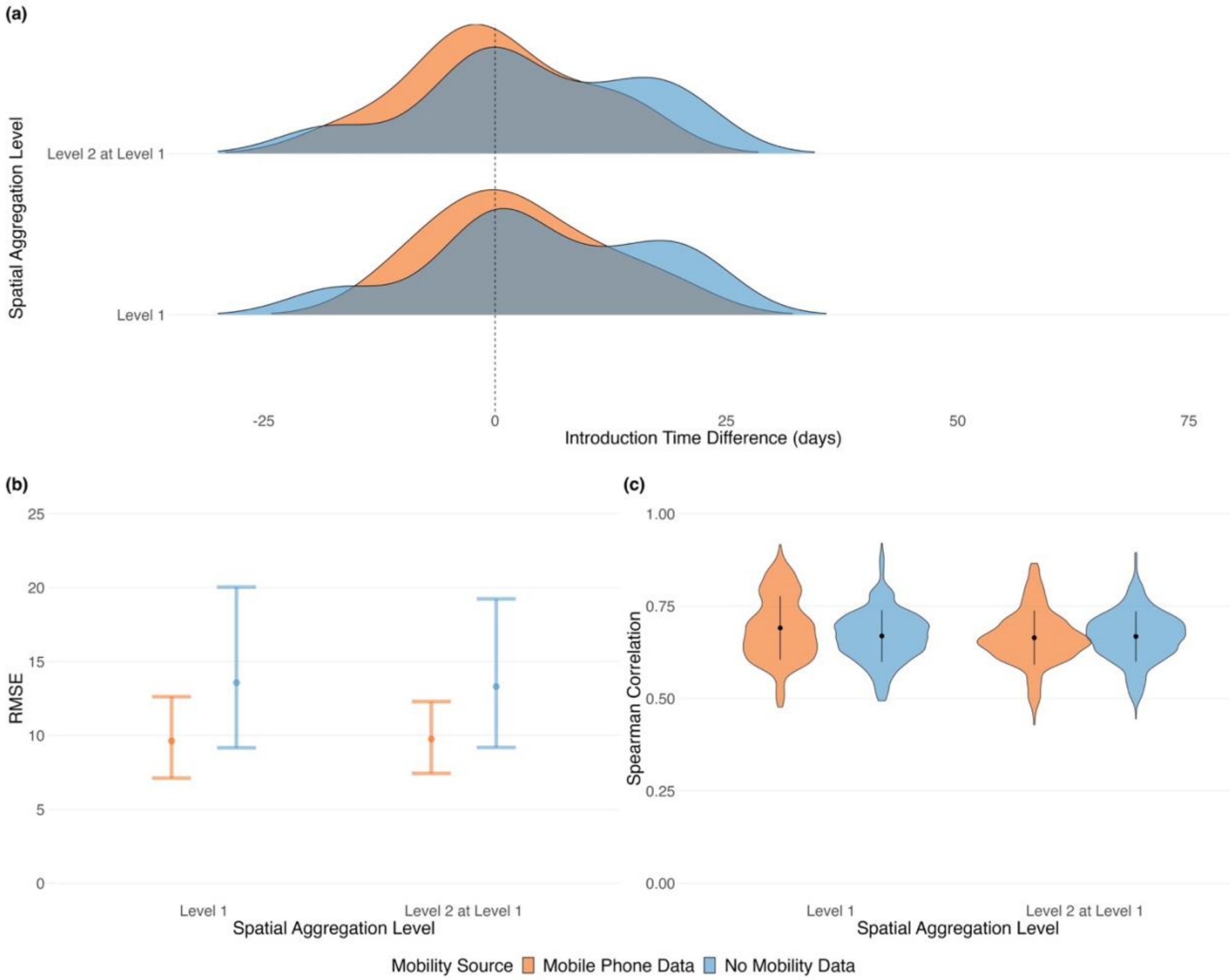

Figure 10. Differences in Spatial Invasion Timing between Model Results and Observed COVID-19 Case Data

(a) Distribution of introduction time differences for models simulated at administrative levels 1 and 2 and with varying assumptions of mobility. Observed introduction times were subtracted from estimated times. Estimated introduction times reflect the average across 1,000 simulations. **(b)** Root mean square error (RMSE) spectrums for each model type. Points display the mean RMSE value across 1,000 simulations and bars depict the 95% range of RMSE values across simulations. **(c)** Spearman rank correlation distributions for each model type from 1,000 simulations. Points represent the mean correlation and lines show one standard deviation above and below the mean.

Similar to patterns observed with spatial invasion, error in epidemic magnitude was smaller in models that incorporated mobility compared to models that did not and in the level 1 mobility model compared to level 2. The administrative level 1 model with mobility had an RMSE of 2,912.28 while the corresponding administrative level 2 model's value was 3,242.74, indicating a slightly closer fit to the COVID-19 data at administrative level 1 (appendix **Figure 18**).

Regardless of the model type, variations in magnitude followed similar trajectories: Differences between observed cases and modeled cases increased as time went on, with most provinces reporting more cases than the model simulated. Interestingly, the Western province began with reporting less cases than the model then switched to recording significantly larger case numbers than the simulation and the Northern province was the single province that consistently reported less cases than the model inferred. Overall, models with mobility were more likely to approximate the observed data than models without mobility for both spatial invasion and epidemic magnitude. Although administrative level 1 models with mobility produced lower error for spatial invasion and epidemic magnitude comparisons than the level 2 model, the differences observed were marginal in relation to the large differences observed in **Figures 7 to 9**.

Discussion

This work highlighted the importance of considering mobility data spatial scale when modeling infectious diseases. Before mobility data were incorporated into models, differences in missingness, trip characteristics, and trip proportions existed between resolution levels and mobility data sources, each with modeling implications. For example, on average, smaller population size units had more external travel than parent units, while units with larger population size had external travel proportions similar to their parent units. This quality of the data was consistent with previous research, where it was found that individuals living in rural areas were more likely to have larger travel radii compared to individuals in urban centers.[15] The phenomenon was replicated in modeling, where introduction locations with less population resulted in larger modeling result differences between mobility spatial levels, due to the higher proportions of external travel in nested units that allowed for disease to spread more rapidly to external parent units. Differences in modeling output were foreshadowed prior to modeling taking place, indicating the usefulness of performing exploratory analyses to gain a better understanding of the critical model input.

Modeling analyses determined that the similarity between model results using finer or coarser mobility data resolutions were dependent on pathogen transmissibility, host duration of infectiousness, as well as the population of the initial seeding location. Models run with varying levels of mobility data had the potential to produce inconsistent results, nonetheless, there should be less concern about large differences in model results if a pathogen is highly infectious, generates long periods of host infectiousness, and is likely to be introduced into an urban area. These findings were supported by previous research concerning the impact of call data record aggregation on infectious disease modeling results in Gahna.[25] Instead of spatial aggregation, the study focused on aggregation methods that summarized the locations visited by users continually moving throughout space. Although different aspects of call data records were aggregated, the study also found that model result discrepancies between aggregation strategies decreased as pathogen transmissibility increased and indicated the importance of introduction location.

For COVID-19 analyses, model results from different mobility spatial resolutions were similar to one another and exhibited marginally improved accuracy from models assuming no mobility. Keeping the results of previous analyses in mind, where model outputs commonly exhibited wide variations across spatial scales, this was surprising. One explanation could be that modeling with observed COVID-19 cases in each district was reflective of the scenarios with an urban introduction location, since the highest number of cases were reported in districts with large population sizes. In the theoretical analyses where an urban introduction was considered, differences between results with varying spatial scales were minimal. Another cause for the similarity of results could be the specific administrative levels compared. Due to case data reporting occurring at the district level, only model results at administrative levels 1 and 2 were simulated. In addition to urban introduction locations producing smaller differences in the results, administrative levels 1 and 2 demonstrated the lowest pairwise variation in results modeled at different spatial levels.

Although COVID-19 models with mobility out-performed models assuming no mobility, exogenous factors might shed light on why there was not a more intense disparity. On April 23, 2021, Sri Lanka introduced restrictions on building capacity, socialization, and ordered the closer of non-essential businesses. On May 10, 2021, an inter-provisional travel ban was issued, and on May 22, 2021, all bus and train services were suspended.[44] Each of these actions had the potential to both physically limit travel and reduce the desire to travel. Given that each of these restrictions occurred during the modeling period, the true level of travel in Sri Lanka during the time was likely somewhere between zero mobility and the level of mobility in the mobile phone data which covered November 2021 to March 2022 when no restrictions were in place. Therefore, the time-period inconsistency between the mobile phone data and the COVID-19 Alpha variant wave could explain why models incorporating mobility were not significantly more accurate than models assuming no mobility.

The results from these analyses are limited in terms of data representativeness and model specifications, however. While mobile phone data had the most complete geographic coverage of the datasets examined, groups such as elderly adults, children, and those of lower socioeconomic status may not be included, which could reduce generalizability of the movement information.[45,46,47] Travel from mobile phone data was also not exact, as cellular towers were used to approximate locations. Censoring was an issue in both the mobile phone data and Facebook data. Trip count information for routes below a threshold were masked, making origin-destination pairs with a small amount of travel and zero travel indistinguishable from one another. In mobile phone data, instances of censoring were assumed to be equivalent to no travel, potentially underestimating trips between geographies. Compared to mobile phone data, the barrier for inclusion in Facebook data was much higher, contributing to data missingness. In order to be included in the Facebook data, individuals needed to own a mobile phone, have access to the Facebook application on their device, and opt-in to location history sharing. Windows of mobility also varied between datasets, where the mobile phone data window was 24-hours, opposed to 8-hours for Facebook data and two weeks prior to dengue infection for travel survey data. Travel window variations likely factored into differences in the nature of travel captured with each dataset, as wider windows allowed for more frequent, longer distance trips.

Assumptions made during the modeling process influenced the validity of these findings as well. Homogeneous mixing within units of the metapopulation was assumed, which was not reflective of reality and changes to this specification have proved to alter model results.[48] The transmissibility and infectiousness duration of the modeled pathogen were held constant throughout each simulation and the proportion of the population susceptible to infection in each unit was uniform, similarly not exhibiting real-world characteristics. The stability of modeling results could also be improved. The metapopulation model run at the finest spatial resolution was computationally intensive, resulting in only 50 simulations per scenario being examined. Increasing the number of simulations may have reduced the variability of epidemic statistics across the spectrums of transmissibility values, durations of infectiousness, and introduction locations explored. Despite the metapopulation model simplifying the complex nature of human behavior, heterogeneities in immunity, and population structure, it is unlikely that the large variations observed between model results at different mobility spatial resolutions would be completely erased if model complexity increased.

In future research, the differences in results between infectious disease models simulated with varying levels of mobility data should be explored beyond mobile phone data, Sri Lanka, and COVID-19. The findings in this work may not be generalizable to all sources of mobility data, to all countries in the world, and to all pathogens, as such, it is essential to investigate if similar conclusions are reached under dissimilar circumstances. Questions of mobility data scale appropriateness remain unanswered. While this work showed that changing the spatial level of mobility data had little improvement on the predictive accuracy of COVID-19 spatial invasion and epidemic magnitude, the result may not be consistent in other scenarios. Research should continue to determine the meaningful scales of human mobility appropriate for disease spatial distribution. Further work should also investigate how the results of movement models, such as gravity models, change with varying levels of mobility data granularity and how those changes impact infectious disease model simulations.

The spatial level of infectious disease models is often seen as an early methodological decision, with little emphasis placed on the significance. This work should encourage researchers to evaluate the scale of their mobility data, and the modeling results produced from them. When infectious disease models are utilized for policy decisions, changing the scale of mobility data may prove as an effective sensitivity analysis to ensure that qualitative findings remain consistent or to understand how inferences are altered if another scale was incorporated. Underpinning the selection of mobility spatial scale should be the mechanistic understanding of how human behavior facilitates disease transmission. As technology continues to become more sophisticated, data on human mobility and the corresponding spatial scales will further increase in granularity, giving way to more estimations of human behavior. Determining which mobility scale best captures relevant behavior and in what contexts the scale is impactful will allow for more accurate modeling results reflective of reality and ultimately meaningful decision-making and public health policy.

Appendix

Table 1. Administrative Level 3 Discrepancies between Geographic Data and Official Sri Lankan Government Distinctions

Geographic Data Divisional Secretariat Name	Official Divisional Secretariat Distinction
Kalthota	Balangoda
Madampagama	Hikkaduwa
Mathurata	Hanguranketa
Nildandahinna	Walapane
Norwood	Ambagamuwa
Rathgama	Hikkaduwa
Thalawakele	Nuwara Eliya
Wanduramba	Baddegama

Table 2. Administrative Level 3 Discrepancies between Mobile Phone Data and Official Sri Lankan Government Distinctions

Mobile Phone Data Divisional Secretariat Name	Official Divisional Secretariat Distinction
Kothmale	Kothmale East
Kothmale	Kothmale West

Table Notes: Each table describes the misalignment in divisional secretariat distinctions between **(a)** geographic data and official government units and **(b)** mobile phone data and official government units. For example, in the geographic data, Hikkaduwa is divided into three units (Hikkaduwa, Madampagama, and Rathgama) while in the mobile phone data, Kothmale is a combination of two units (Kothmale East and Kothmale West). For all analyses, data sources were aligned to the 330 divisional secretariate distinctions in the mobile phone data.

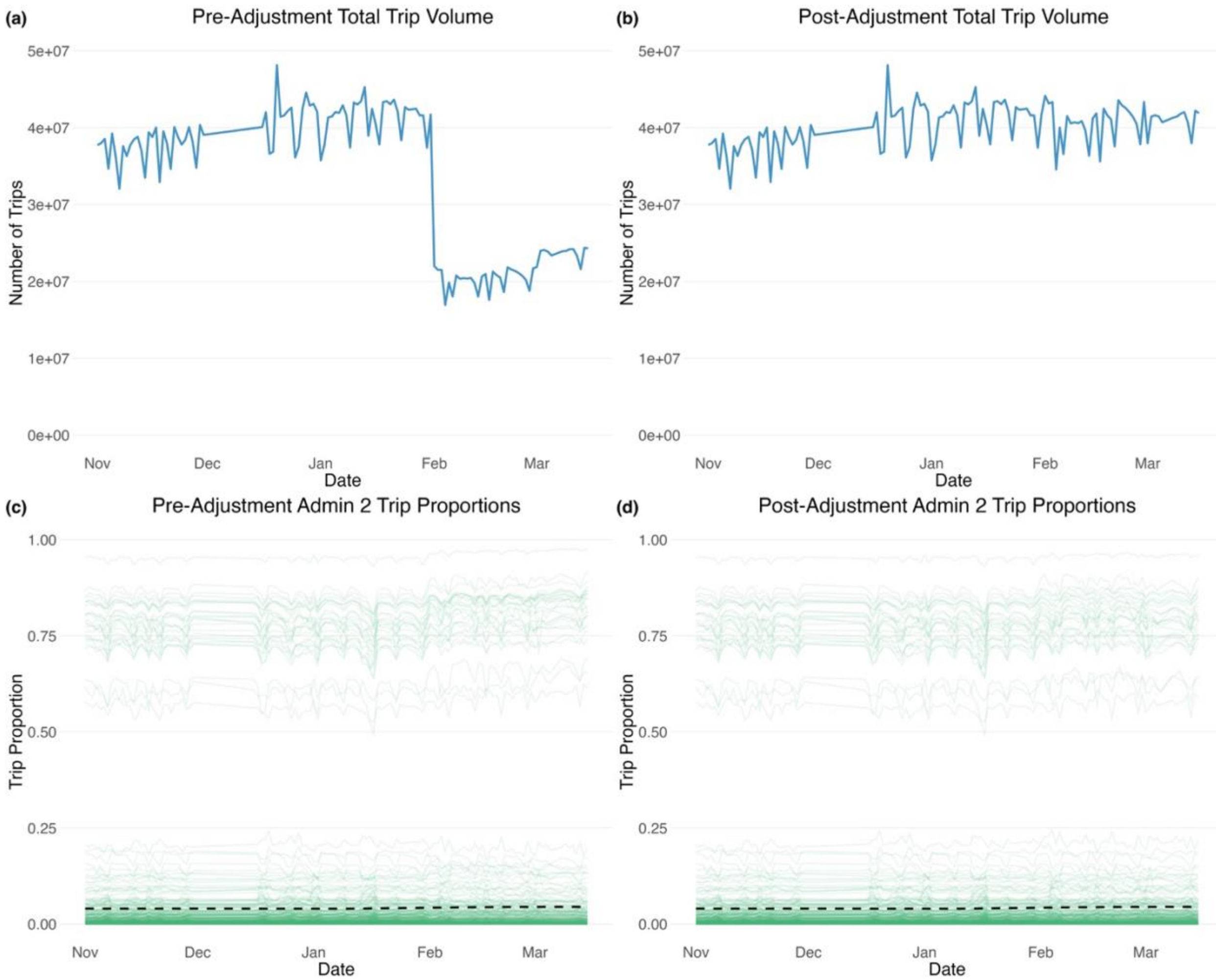

Figure 11. Mobile Phone Data Adjustment by Total Volume and Proportion

(a) Pre-adjustment total trip volume for mobile phone data. **(b)** Total trip volume for mobile phone data following volume adjustment, scaling February and March volume to match January volume. **(c)** Pre-volume adjustment trip proportion for each origin-destination pair at administrative level 2. **(d)** Post-volume adjustment trip proportion for each origin-destination pair at administrative level 2. The black dashed line in both **(c)** and **(d)** depicts locally estimated scatterplot smoothing (LOESS) average trip proportion across time.

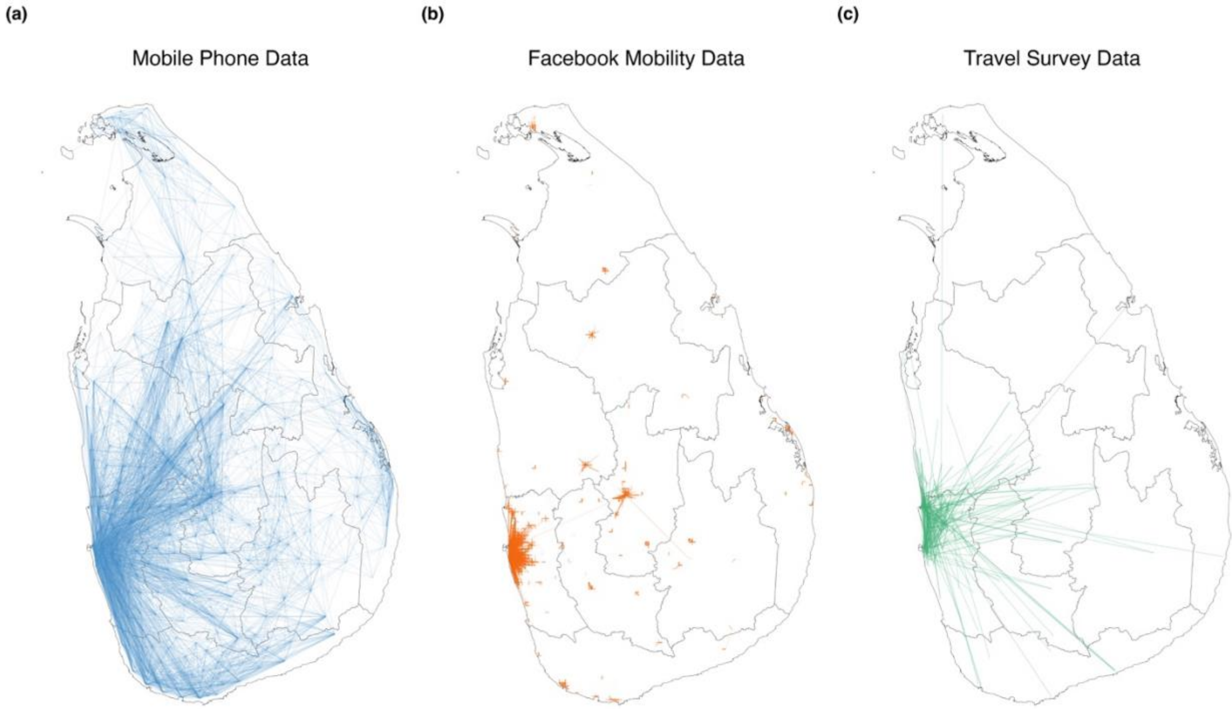

Figure 12. Fine-Scale Mobility Maps by Mobility Data Source

For **(a)** mobile phone data at administrative level 3, **(b)** Facebook mobility data at the tile level, and **(c)** dengue travel survey data at the location level. Each map displays the finest scale mobility for the data source. Mobile phone trajectories represent the top 10% of routes by daily average trip volume while Facebook mobility and travel survey maps show all routes present in each dataset.

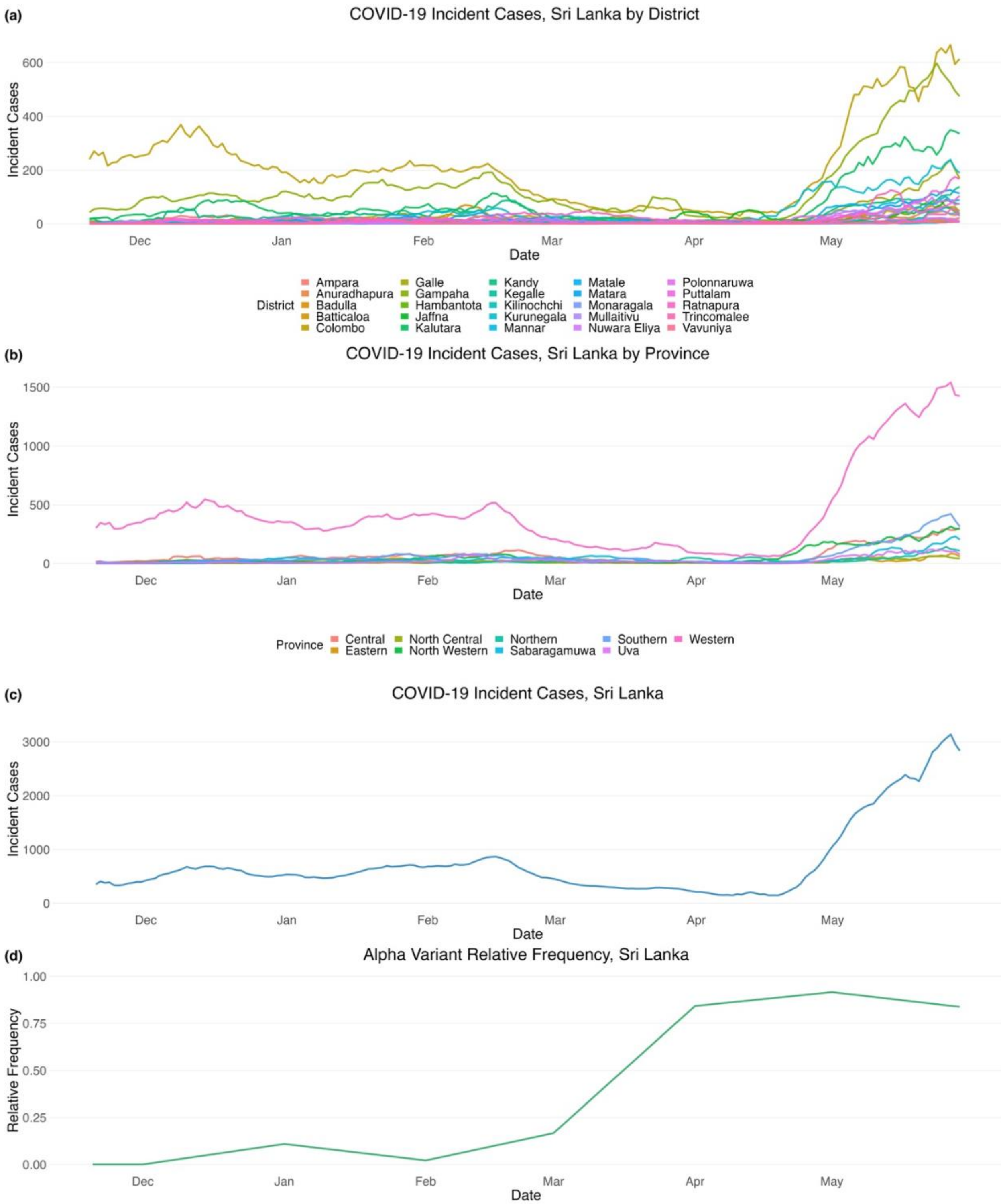

Figure 13. COVID-19 in Sri Lanka from November 14, 2020 to May 29, 2021

(a) 7-day rolling average incident COVID-19 cases by Sri Lanka district. **(b)** 7-day rolling average incident COVID-19 cases by Sri Lanka province. **(c)** 7-day rolling average incident COVID-19 cases for Sri Lanka. **(d)** Relative frequency of the COVID-19 Alpha variant in Sri Lanka, from GISAID.[49]

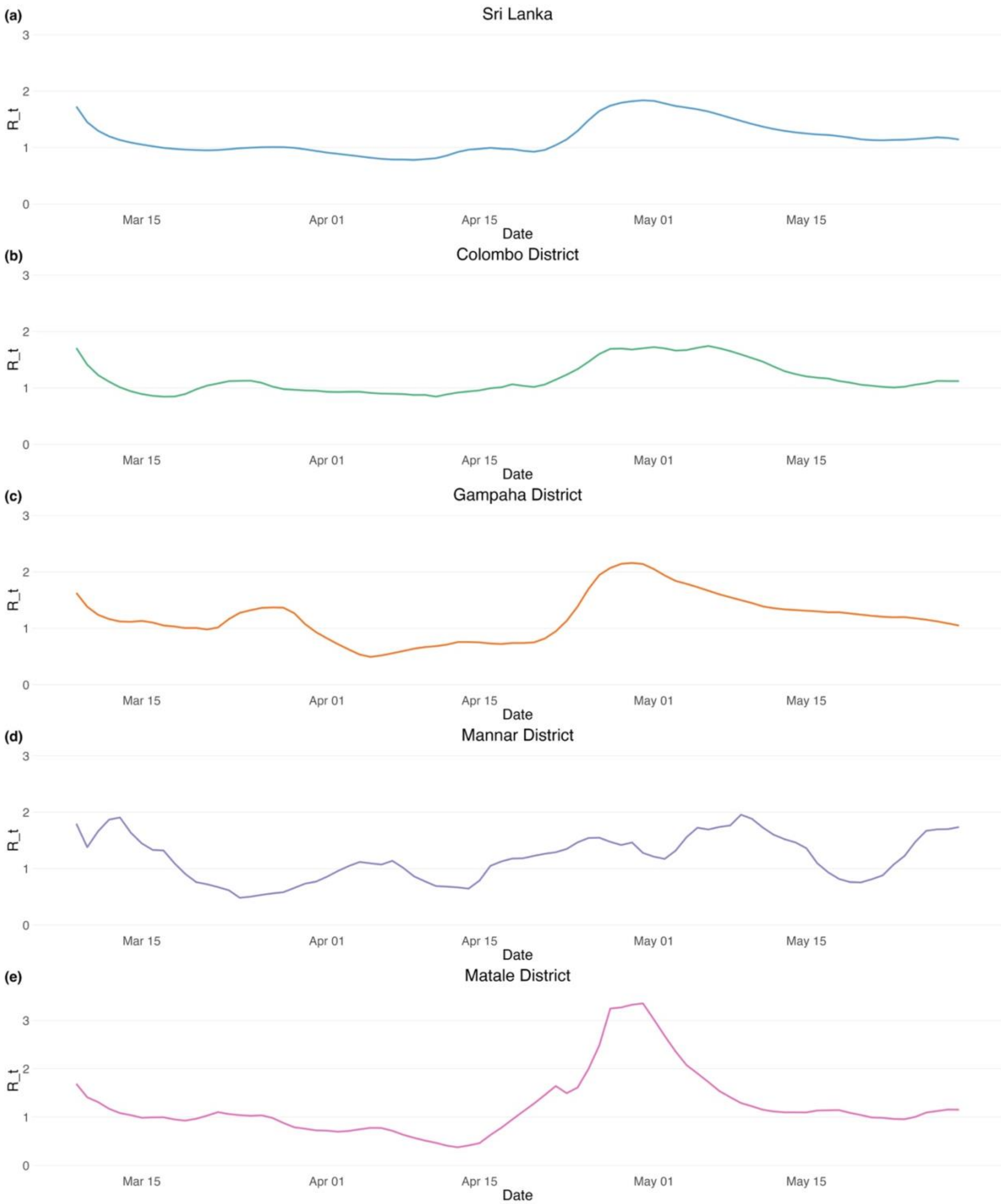

Figure 14. Effective Reproduction Number (R_t) by Location from March 1 – May 29, 2021

For **(a)** Sri Lanka, **(b)** Colombo District, **(c)** Gampaha District, **(d)** Mannar District, and **(e)** Matale District. Effective reproduction number (R_t) was calculated with the EpiEstim package in R, using the parametric serial interval method assuming a mean serial interval of 5.2 with a standard deviation of 4.75, each informed by the literature.[50,51]

Table 3. Linear Relationship between Logarithm Trip Count and Distance by Mobility Data Source and Administrative Level

Data Source	Linear Relationship between Trip Count and Distance		
	Admin Level 1	Admin Level 2	Admin Level 3
Mobile Phone	-0.028	-0.024	-0.011
Facebook Mobility	-0.046	-0.105	-0.115
Travel Survey	-0.009	-0.007	-0.005

Table 4. Mobility Descriptive Statistics by Mobility Data Source and Administrative Level

Admin Level	Data Source	Total Daily Average Trips*	Maximum Daily Average Trips*	Average Trip Distance (km)	Maximum Trip Distance (km)
1	Mobile Phone	40,166,011	14,713,697	9.230	327.139
	Facebook Mobility	772,406	428,603	0.230	257.746
	Travel Survey	658	461	19.568	257.746
2	Mobile Phone	40,167,885	6,439,044	11.642	397.359
	Facebook Mobility	627,359	214,753	0.274	95.302
	Travel Survey	658	360	23.429	285.577
3	Mobile Phone	40,987,358	568,713	17.807	429.791
	Facebook Mobility	630,450	40,499	0.500	93.612
	Travel Survey	658	88	21.621	298.346

Table Notes: Descriptive statistics were calculated for each mobility data source and administrative level. **(a)** Depicts the linear relationship between logarithm trip count and distance displayed in **Figure 2**. **(b)** Lists total daily average trips, maximum daily average trips, average trip distance, and maximum trip distance.
(*) Daily average trip volume across time was only able to be calculated for mobile phone and Facebook mobility data due to the longitudinal nature of each source. For travel survey data, statistics are calculated from the number of trips in a two-week period prior to dengue infection and do not represent a daily average.

Table 5. Missing Data Univariate Regression Results by Explanatory Variable, Mobility Data Source, and Administrative Level

Univariate Explanatory Variable	Data Source	Administrative Level	Coefficient	P-Value
Origin Population	Mobile Phone	3	-3.804	<2.00E-16***
	Facebook	1	-0.092	0.012*
		2	-0.049	0.026*
		3	-0.960	<2.00E-16***
	Travel Survey	1	-0.091	0.00039***
		2	-0.176	<2.00E-16***
		3	-1.476	<2.00E-16***
Destination Population	Mobile Phone	3	-4.907	<2.00E-16***
	Facebook	1	-0.011	0.49
		2	-0.064	0.002**
		3	-0.973	<2.00E-16***
	Travel Survey	1	-0.049	<2.00E-16***
		2	-0.142	0.004**
		3	-1.600	<2.00E-16***
Trip Distance	Mobile Phone	3	0.024	<2.00E-16***
	Facebook	1	3.420	9.72E-06***
		2	8.926	4.02E-11***
		3	28.687	<2.00E-16***
	Travel Survey	1	1.302	0.004**
		2	1.548	1.09E-08***
		3	3.548	<2.00E-16***

Significance: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table Notes: Population variables are scaled such that coefficients represent change in log odds per change in 100,000 people. Distance variable is scaled so coefficients represent change in log odds per change in 100 km trip distance. Administrative levels 1 and 2 are not present for mobile phone data since no missingness was observed in those datasets.

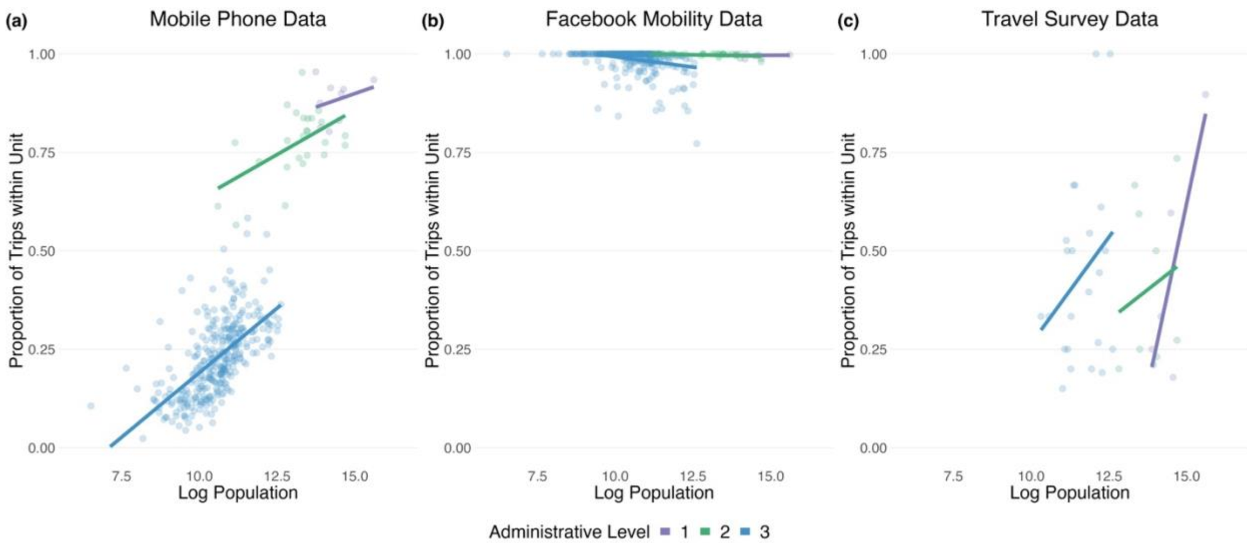

Figure 15. Internal Trip Proportion Distribution and Population by Mobility Dataset and Administrative Level

(a) Relationship between internal trip proportion and logarithm unit population for mobile phone data administrative levels 1, 2, and 3. **(b)** Relationship between internal trip proportion and logarithm unit population for Facebook mobility data administrative levels 1, 2, and 3. **(c)** Relationship between internal trip proportion and logarithm unit population for travel survey data administrative levels 1, 2, and 3.

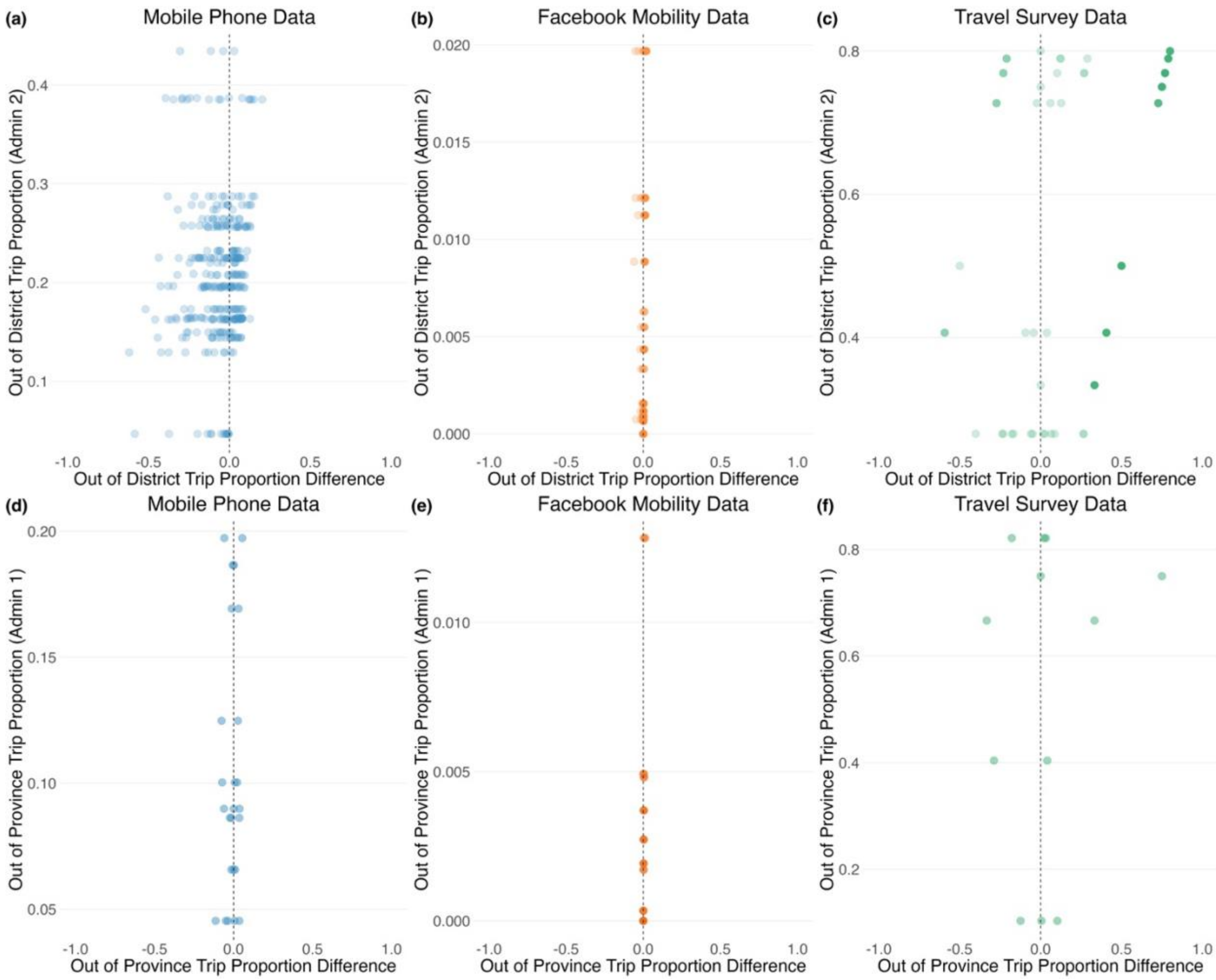

Figure 16. Differences in External Travel between Administrative Levels 2 and 3 and Administrative Levels 1 and 2

(a), (b), and (c) Difference in out of district travel between administrative 2 units and nested administrative level 3 units by administrative level 2 out of district trip proportion for mobile phone, Facebook, and travel survey data. Trip proportion differences were calculated by subtracting administrative level 3 out of district trip proportion from administrative level 2 out of district trip proportion. **(d), (e), and (f)** Difference in out of province travel between administrative 1 units and nested administrative level 2 units by administrative level 1 out of province trip proportion for mobile phone, Facebook, and travel survey data. Trip proportion differences were calculated by subtracting administrative level 2 out of province trip proportion from administrative level 1 out of province trip proportion.

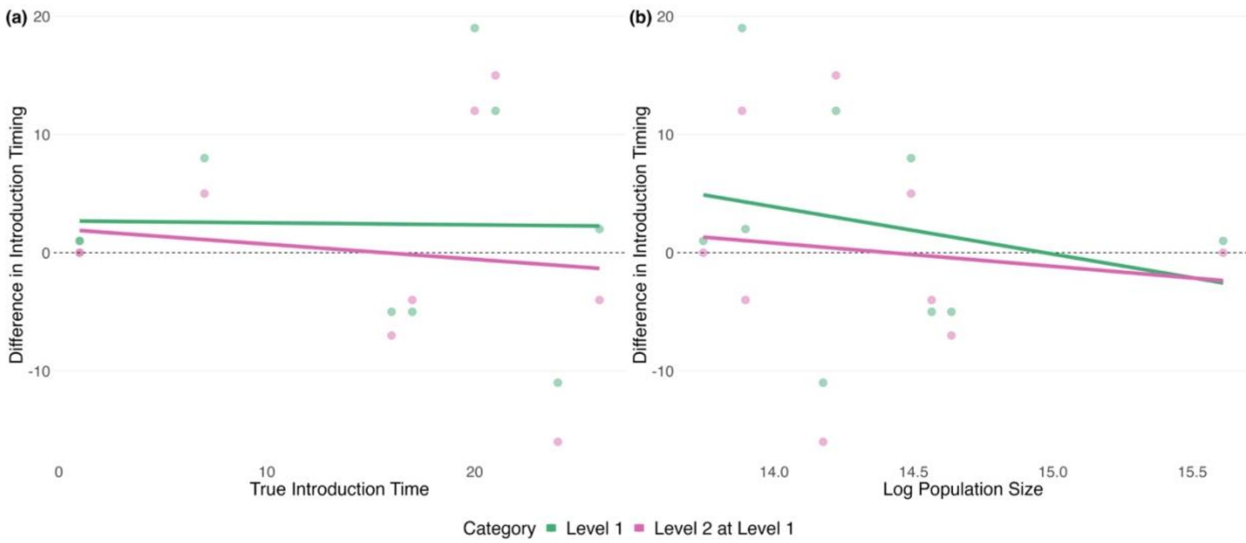

Figure 17. Differences in Spatial Invasion Timing between Model Results and Observed COVID-19 Case Data

(a) The difference in spatial invasion timing between model results and observed COVID-19 case data for models with administrative level 1 mobility data and administrative level 2 mobility data aggregated to the province-level by true introduction time. **(b)** The difference in spatial invasion timing between model results and observed COVID-19 case data for models with administrative level 1 mobility data and administrative level 2 mobility data aggregated to the province-level by logarithm of unit population size. Differences were calculated by subtracting the true COVID-19 introduction time to the province from the modeled introduction time, with each point representing a province.

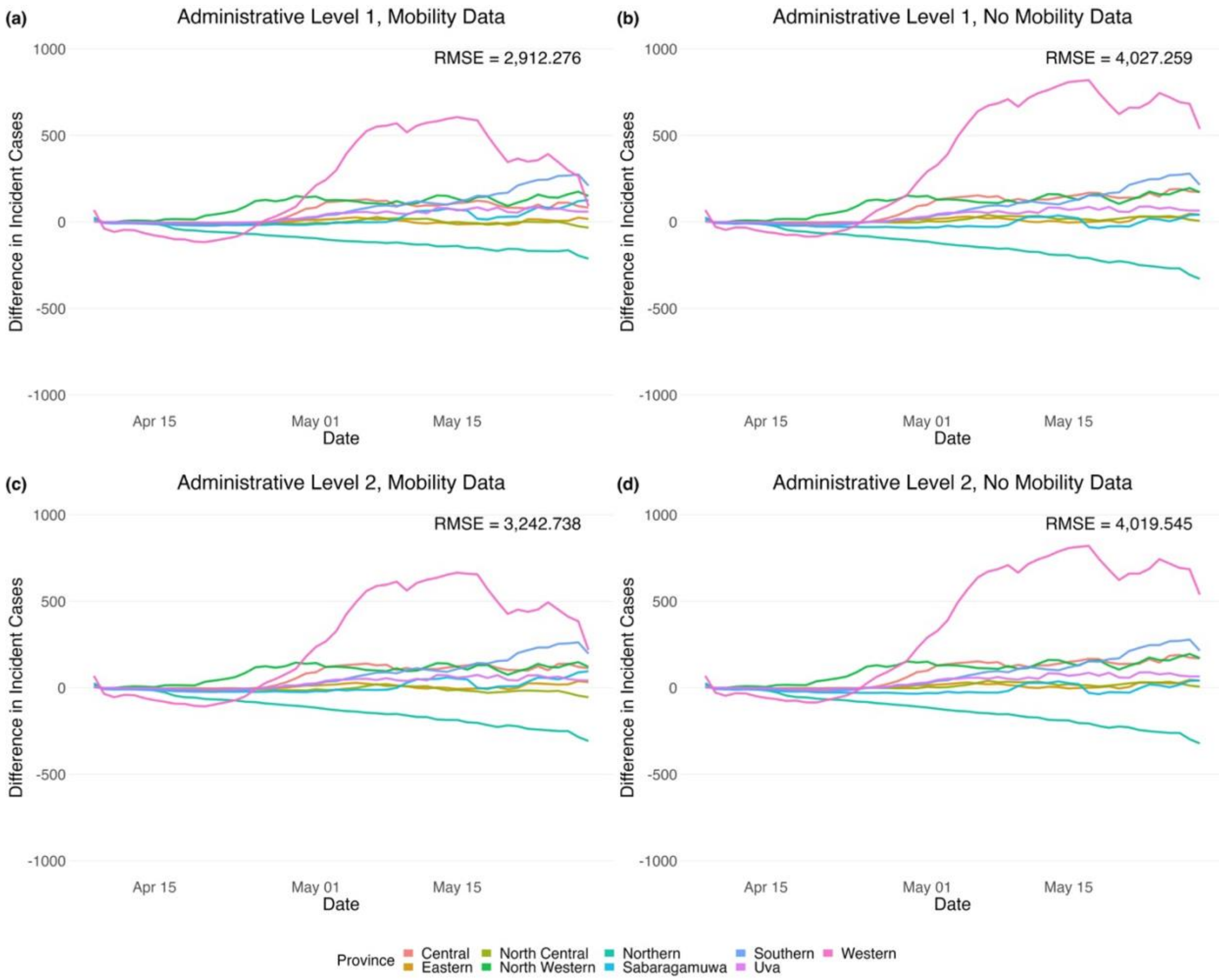

Figure 18. Differences in Epidemic Magnitude between Model Results and Observed COVID-19 Case Data

(a) Differences between province-level reported COVID-19 cases and predicted cases from the model simulated with administrative level 1 mobility data. **(b)** Differences between province-level reported COVID-19 cases and predicted cases from the model simulated at administrative level 1 assuming no mobility. **(c)** Differences between province-level reported COVID-19 cases and predicted cases from the model simulated with administrative level 2 mobility data. **(d)** Differences between province-level reported COVID-19 cases and predicted cases from the model simulated at administrative level 2 assuming no mobility. Differences were calculated by subtracting predicted cases from observed COVID-19 cases.

Bibliography

1. Davis JT, Chinazzi M, Perra N, et al. Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave. *Nature*. 2021;600(7887):127-132. doi:10.1038/s41586-021-04130-w

2. Jo Y, Sung H. Impact of pre-pandemic travel mobility patterns on the spatial diffusion of COVID-19 in South Korea. *J Transp Health*. 2022;26:101479. doi:10.1016/j.jth.2022.101479

3. Engebretsen S, Engø-Monsen K, Aleem MA, Gurley ES, Frigessi A, de Blasio BF. Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: the case of Bangladesh. *J R Soc Interface*. 2020;17(167):20190809. doi:10.1098/rsif.2019.0809

4. Charu V, Zeger S, Gog J, et al. Human mobility and the spatial transmission of influenza in the United States. *PLOS Comput Biol*. 2017;13(2):e1005382. doi:10.1371/journal.pcbi.1005382

5. Wesolowski A, Eagle N, Tatem AJ, et al. Quantifying the impact of human mobility on malaria. *Science*. 2012;338(6104):267-270. doi:10.1126/science.1223467

6. Wesolowski A, Qureshi T, Boni MF, et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proc Natl Acad Sci*. 2015;112(38):11887-11892. doi:10.1073/pnas.1504964112

7. Kraemer MUG, Faria NR, Reiner RC, et al. Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015-16: a modelling study. *Lancet Infect Dis*. 2017;17(3):330-338. doi:10.1016/S1473-3099(16)30513-8

8. Bengtsson L, Gaudart J, Lu X, et al. Using mobile phone data to predict the spatial spread of cholera. *Sci Rep*. 2015;5:8923. doi:10.1038/srep08923

9. Wesolowski A, Winter A, Tatem AJ, et al. Measles outbreak risk in Pakistan: exploring the potential of combining vaccination coverage and incidence data with novel data-streams to strengthen control. *Epidemiol Infect*. 2018;146(12):1575-1583. doi:10.1017/S0950268818001449

10. Tatem AJ. Mapping population and pathogen movements. *Int Health*. 2014;6(1):5-11. doi:10.1093/inthealth/ihu006

11. Lai S, Farnham A, Ruktanonchai NW, Tatem AJ. Measuring mobility, disease connectivity and individual risk: a review of using mobile phone data and mHealth for travel medicine. *J Travel Med*. 2019;26(3):taz019. doi:10.1093/jtm/taz019

12. Vazquez-Prokopec GM, Bisanzio D, Stoddard ST, et al. Using GPS Technology to Quantify Human Mobility, Dynamic Contacts and Infectious Disease Dynamics in a Resource-Poor Urban Environment. *PLoS ONE*. 2013;8(4):e58802. doi:10.1371/journal.pone.0058802

13. Wesolowski A, Buckee CO, Engø-Monsen K, Metcalf CJE. Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data. *J Infect Dis*. 2016;214(Suppl 4):S414-S420. doi:10.1093/infdis/jiw273

14. Maas P. Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM; 2019:3173-3173. doi:10.1145/3292500.3340412

15. Alessandretti L, Aslak U, Lehmann S. The scales of human mobility. *Nature*. 2020;587(7834):402-407. doi:10.1038/s41586-020-2909-1

16. Iwasa Y, Andreasen V, Levin S. Aggregation in model ecosystems. I. Perfect aggregation. *Ecol Model*. 1987;37(3):287-302. doi:10.1016/0304-3800(87)90030-5

17. Riley S, Eames K, Isham V, Mollison D, Trapman P. Five challenges for spatial epidemic models. *Epidemics*. 2015;10:68-71. doi:10.1016/j.epidem.2014.07.001

18. Borchering RK, Viboud C, Howerton E, et al. Modeling of Future COVID-19 Cases, Hospitalizations, and Deaths, by Vaccination Rates and Nonpharmaceutical Intervention Scenarios — United States, April–September 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(19):719-724. doi:10.15585/mmwr.mm7019e3

19. Spicknall IH. Modeling the Impact of Sexual Networks in the Transmission of Monkeypox virus Among Gay, Bisexual, and Other Men Who Have Sex With Men — United States, 2022. *MMWR Morb Mortal Wkly Rep*. 2022;71. doi:10.15585/mmwr.mm7135e2

20. Meltzer MI. Modeling in Real Time During the Ebola Response. *MMWR Suppl*. 2016;65. doi:10.15585/mmwr.su6503a12

21. Bauch CT, Lloyd-Smith JO, Coffee MP, Galvani AP. Dynamically Modeling SARS and Other Newly Emerging Respiratory Illnesses: Past, Present, and Future. *Epidemiology*. 2005;16(6):791. doi:10.1097/01.ede.0000181633.80269.4c

22. Wesolowski A, Metcalf CJE, Eagle N, et al. Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proc Natl Acad Sci U S A*. 2015;112(35):11114-11119. doi:10.1073/pnas.1423542112

23. Truscott J, Ferguson NM. Evaluating the Adequacy of Gravity Models as a Description of Human Mobility for Epidemic Modelling. *PLOS Comput Biol*. 2012;8(10):e1002699. doi:10.1371/journal.pcbi.1002699

24. Ramiadantsoa T, Metcalf CJE, Raherinandrasana AH, et al. Existing human mobility data sources poorly predicted the spatial spread of SARS-CoV-2 in Madagascar. *Epidemics*. 2022;38:100534. doi:10.1016/j.epidem.2021.100534

25. Gibbs H, Musah A, Seidu O, et al. Call detail record aggregation methodology impacts infectious disease models informed by human mobility. Published online January 28, 2023:2023.01.25.23285005. doi:10.1101/2023.01.25.23285005

26. Sri Lanka - Subnational Administrative Boundaries - Humanitarian Data Exchange. Accessed February 2, 2023. https://data.humdata.org/dataset/geoboundaries-admin-boundaries-for-sri-lanka

27. Grama Niladhari Administration Division. Ministry of Public Administration and Management. Published September 14, 2016. Accessed March 13, 2023. https://web.archive.org/web/20160914121553/http://www.pubad.gov.lk/web/index.php?option=com_content&view=category&id=19&layout=blog&Itemid=65&lang=en

28. Population Counts. WorldPop Hub. Accessed February 2, 2023. https://hub.worldpop.org/geodata/listing?id=79

29. World Population Prospects 2022. United Nations, Department of Economic and Social Affairs - Population Division. Accessed March 22, 2023. https://population.un.org/wpp/Download/Standard/MostUsed/

30. Esri Inc. ArcGIS Pro. Published online 2021. https://pro.arcgis.com/en/pro-app/2.9/get-started/download-arcgis-pro.htm

31. Dialog 2021 Annual Report. Dialog Axiata PLC. Published 2021. Accessed February 3, 2023. https://s3.ap-southeast-1.amazonaws.com/dlg.dialog.lk/s3fs-public/2022-06/ro_annual-report-2021.pdf

32. Facebook Disaster Maps: Methodology - Meta Research. Meta Research. Accessed March 22, 2023. https://research.facebook.com/blog/2017/6/facebook-disaster-maps-methodology/

33. Brundritt R. Bing Maps Tile System - Bing Maps. Microsoft. Published June 8, 2022. Accessed February 8, 2023. https://learn.microsoft.com/en-us/bingmaps/articles/bing-maps-tile-system

34. Geocoding API. Google Maps Platform. Accessed February 15, 2023. https://developers.google.com/maps/documentation/geocoding

35. COVID-19 Situation Report. Epidemiology Unit, Ministry of Health, Sri Lanka. Accessed February 8, 2023. https://www.epid.gov.lk/web/index.php?option=com_content&view=article&id=238&Itemid=519&lang=en

36. Citron DT, Guerra CA, Dolgert AJ, et al. Comparing metapopulation dynamics of infectious diseases under different models of human movement. *Proc Natl Acad Sci*. 2021;118(18):e2007488118. doi:10.1073/pnas.2007488118

37. Xia Y, Bjørnstad ON, Grenfell BT. Measles Metapopulation Dynamics: A Gravity Model for Epidemiological Coupling and Dynamics. *Am Nat*. 2004;164(2):267-281. doi:10.1086/422341

38. Cori A, Ferguson NM, Fraser C, et al. EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves. Published online January 7, 2021. Accessed March 28, 2023. https://CRAN.R-project.org/package=EpiEstim

39. Walsh KA, Spillane S, Comber L, et al. The duration of infectiousness of individuals infected with SARS-CoV-2. *J Infect*. 2020;81(6):847-856. doi:10.1016/j.jinf.2020.10.009

40. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *Lancet Microbe*. 2021;2(1):e13-e22. doi:10.1016/S2666-5247(20)30172-5

41. Kissler SM, Fauver JR, Mack C, et al. Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. *PLOS Biol*. 2021;19(7):e3001333. doi:10.1371/journal.pbio.3001333

42. R Core Team. R: A Language and Environment for Statistical Computing. Published online 2022. https://www.R-project.org/

43. HopkinsIDD/spatial-resolution-sri-lanka: Ronan Corgel's ScM thesis project. Accessed April 5, 2023. https://github.com/HopkinsIDD/spatial-resolution-sri-lanka

44. COVID-19 Pandemic - Sri Lanka. A3M Global Monitoring. Accessed March 30, 2023. https://global-monitoring.com/gm/page/events/epidemic-0002015.fTDtGCxti2qN.html?lang=en

45. Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO. The impact of biases in mobile phone ownership on estimates of human mobility. *J R Soc Interface*. 2013;10(81):20120986. doi:10.1098/rsif.2012.0986

46. Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO. Heterogeneous mobile phone ownership and usage patterns in Kenya. *PloS One*. 2012;7(4):e35319. doi:10.1371/journal.pone.0035319

47. Grantz KH, Meredith HR, Cummings DAT, et al. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat Commun*. 2020;11:4961. doi:10.1038/s41467-020-18190-5

48. Aleta A, Ferraz de Arruda G, Moreno Y. Data-driven contact structures: From homogeneous mixing to multilayer networks. *PLoS Comput Biol*. 2020;16(7):e1008035. doi:10.1371/journal.pcbi.1008035

49. hCoV-19 Variants Dashboard. GISAID. Published 2021. Accessed March 28, 2023. https://gisaid.org/hcov-19-variants-dashboard/

50. Alene M, Yismaw L, Assemie MA, Ketema DB, Gietaneh W, Birhan TY. Serial interval and incubation period of COVID-19: a systematic review and meta-analysis. *BMC Infect Dis*. 2021;21(1):257. doi:10.1186/s12879-021-05950-x

51. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerg Infect Dis*. 2020;26(6):1341-1343. doi:10.3201/eid2606.200357