

A systematic review of using population-level human mobility data to understand SARS-CoV-2 transmission

Received: 14 July 2024

Accepted: 25 November 2024

Published online: 03 December 2024

 Check for updates

Natalya Kostandova^{1,19} , Catherine Schluth^{1,19}, Rohan Arambepola¹, Fatumah Atuhaire², Sophie Bérubé¹, Taylor Chin³, Eimear Cleary², Oscar Cortes-Azuero⁴, Bernardo García-Carreras⁵, Kyra H. Grantz¹, Matt D. T. Hitchings^{6,7}, Angkana T. Huang^{4,5}, Nishant Kishore⁸, Shengjie Lai², Sophie L. Larsen⁹, Stacie Loiselet¹, Pamela Martinez^{10,11}, Hannah R. Meredith¹, Ritika Purbey¹², Tanjona Ramiadantsoa^{13,14}, Jonathan Read¹⁵, Benjamin L. Rice¹⁶, Lori Rosman¹⁷, Nick Ruktanonchai^{2,18}, Henrik Salje¹⁴, Kathryn L. Schaber¹, Andrew J. Tatem², Jasmine Wang¹, Derek A. T. Cummings^{5,7,20} & Amy Wesolowski^{1,20}

The emergence of SARS-CoV-2 into a highly susceptible global population was primarily driven by human mobility-induced introduction events. Especially in the early stages, understanding mobility was vital to mitigating the pandemic prior to widespread vaccine availability. We conducted a systematic review of studies published from January 1, 2020, to May 9, 2021, that used population-level human mobility data to understand SARS-CoV-2 transmission. Of the 5505 papers with abstracts screened, 232 were included in the analysis. These papers focused on a range of specific questions but were dominated by analyses focusing on the USA and China. The majority included mobile phone data, followed by Google Community Mobility Reports, and few included any adjustments to account for potential biases in population sampling processes. There was no clear relationship between methods used to integrate mobility and SARS-CoV-2 data and goals of analysis. When considering papers focused only on the estimation of the effective reproductive number within the US, there was no clear relationship identified between this measure and changes in mobility patterns. Our findings underscore the need for standardized, systematic ways to identify the source of mobility data, select an appropriate approach to using it in analysis, and reporting.

Prior to the widespread availability of vaccination, mitigation measures for curtailing transmission of SARS-CoV-2 primarily relied on non-pharmaceutical interventions (NPIs) that aimed to reduce the contact between susceptible and infected individuals^{1–4}. NPIs included business and school closures, stay-at-home orders, limited gatherings, and travel restrictions, although their implementation and

enforcement varied geographically and temporally^{5–7}. As most NPIs relied on limiting contact through restrictions on mobility, they resulted in an unprecedented change to human travel patterns globally. While human travel has historically been studied in relation to infectious disease transmission⁸, the start of the COVID-19 pandemic in late 2019 led to a spur of research and analyses on human travel for

A full list of affiliations appears at the end of the paper. ✉ e-mail: nkostan1@jh.edu

both scientific insight and to help guide public health policymaking. In contrast to other examples of human mobility driving disease transmission, the COVID-19 pandemic was one of the first instances in which data have allowed for the quantification of these behaviors in near real-time to inform our understanding of outbreak dynamics.

Data on human mobility became widely available to researchers and policymakers. While some data sources, such as the census, traffic data, and public transport timetables, have been used as traditional sources of mobility information going back as far as a century, more recently novel datasets including social media, internet check-in, and mobile phone call detail records emerged as alternative sources of mobility data^{9–16}. Often, these data were integrated into analyses of SARS-CoV-2 epidemiology to understand changes in mobility or contact patterns and their implications for disease transmission, and to inform policy decisions. However, to date, there has been limited evaluation of what datasets were used, how they were processed and analyzed, and, importantly, how these data informed our understanding of the intersection between SARS-CoV-2 transmission and human behavior. Further, while these datasets and approaches have provided an opportunity to quantify the relationship between transmission, behavior, and policies in place, the utility of these approaches as the COVID-19 pandemic progressed, and for other pathogens, remains unclear. This study was not the first to analyze the use of mobility data in understanding the COVID-19 pandemic. Zhang and colleagues (2022)¹⁷ reviewed literature investigating associations between mobility and COVID-19 published before September 2020. They found an astonishing variety of statistical, mechanistic, and arithmetic models used in the papers; within the 47 articles included in the review, 29 different models were used. Given the review's focus on the first year of the COVID-19 pandemic, we anticipated that studies published after September 2020 would have shifted in research objectives, geographic focus, and methods used to treat and integrate mobility data.

Therefore, we analyzed scientific papers published during the early stage (December 2019–May 2021) of the COVID-19 pandemic to investigate how and what type of mobility data were used to conduct analyses of SARS-CoV-2 transmission. We focused only on papers analyzing population-level human mobility data (at least 1000 individuals) to understand the transmission of SARS-CoV-2. We investigated the objectives of the studies carried out, the type of data used, how these data were processed and integrated into analyses of SARS-CoV-2 transmission, and the results presented in the studies. Finally, we highlighted areas where these approaches were able to add valuable insight but also identified remaining gaps in our understanding of human mobility and the transmission of acute respiratory infectious diseases globally.

Results

Screening processing and final papers for review

A total of 5505 abstracts were screened. Abstracts were further categorized based on their topic. Of the 5138 abstracts excluded, 1574 abstracts were focused on SARS-CoV-2 but did not use mobility data, or were a case report; 1372 were not about SARS-CoV-2; 362 were either a bad return (e.g., no paper at the link), or identified as a duplicate; 753 were categorized as not research and primarily focused on opinion pieces/perspectives/reports on guidelines; and 1077 were an erratum or correction to another paper, or did not include an abstract (Supplementary Fig. 1).

Of 367 papers assessed for eligibility during full-text review, 135 were excluded, as they did not include mobility data and/or included data on less than 1000 individuals, or were published outside the period of interest, leaving 232 papers for analysis (Supplementary Fig. 1). A full bibliography of these papers is provided in Supplementary Methods 4.

Section 0: Basic information on published papers

The included papers were published between January 2020 and May 2021 in 112 different journals. Journal areas covered a wide range of fields, including general scientific audiences, engineering, epidemiology, biology, mathematics, and statistics. The vast majority of first authors were affiliated with academic institutions; 91 (39%) were affiliated with epidemiology, medicine, or public health departments of academia, 68 (29%) were affiliated with computer science, engineering, or geography departments, and 31 (13%) were affiliated with business, economics, or social science departments (Supplementary Table 1). Less commonly, first authors were affiliated with other academic departments, such as biology or ecology (21, 9%); statistics, biostatistics, or mathematics (20, 9%); or other (9, 4%). Only 11 first authors were affiliated with non-academic, non-government agencies (7 agencies not in public health, and 4 in public health), and only 6 had government agency affiliations (5 in public health and 1 not in public health). Four first authors were affiliated with healthcare institutions such as medical institutions, clinics, or hospitals. Articles published during the first five months of the pandemic included first authors with academic affiliations only (Supplementary Fig. 2A). The distribution of senior author affiliations was similarly dominated by academia (Supplementary Table 2). Only 11% of papers included any author affiliated with a public health agency; these papers were published throughout the period of analysis (Supplementary Fig. 3).

Section 1: Area of focus

Most manuscripts focused on a single country at a national or subnational level (162) or within a broader global context (15). Of the remaining papers, these tended to focus on multiple countries (29), with very few purely theoretical (4) without a specific location(s) listed (Fig. 1A). Papers with a global focus were not published within the first five months of the time frame specified in our analyses (May 2020, Fig. 1A); however, the change in area of focus over time was not statistically significant (Table 1). For papers focused on a single country, analyses were dominated by two countries, the United States of America (USA) (52, 32%) and China (47, 29%) (Fig. 1B), accounting for over half of single-country focused papers, and nearly half of all papers included. The next most common locations were in Europe or in other high-income countries (Italy = 9, UK = 6, Japan = 7, Canada = 4, Australia = 3), and few in low or middle-income countries (LMIC, as defined by the World Bank¹⁸) (excluding China). Outside the US, in the Americas, only Brazil had five or more papers (5), and LMIC countries only represented 4% of single-country focused analyses. Unsurprisingly, European countries were more likely to be included in papers with a multinational focus, with at least one European country included in 22 of 29 papers, 14 of which focused only on European countries. Only 11 multinational papers focused on Asian countries (excluding China), four papers on countries in Africa, five on countries in North America, two on countries in Oceania, and one on countries in South America. Of papers that listed specific dates and an explicit time frame for their analyses ($n = 196$), analyses covered between less than one month and two years, with most covering a few months (median = 3 months, mean = 3.8 months), with duration of analysis increasing over time (Supplementary Fig. 4). The time frame tended to focus on the mid/end of March 2020 when the first restrictions and NPIs were implemented. Few papers investigated further than mid-2020 as the end period for the analysis.

The largest proportion of papers aimed to quantify the association between mobility and transmission of SARS-CoV-2 (81, 37%), with some conducted to assess changes in mobility during this period (37, 17%) (Supplementary Table 3). Several papers focused on specific non-pharmaceutical interventions, including lockdowns (74), travel restrictions (50), quarantining (6), and masking (5). A smaller proportion of papers was devoted to introduction events, including 41 on

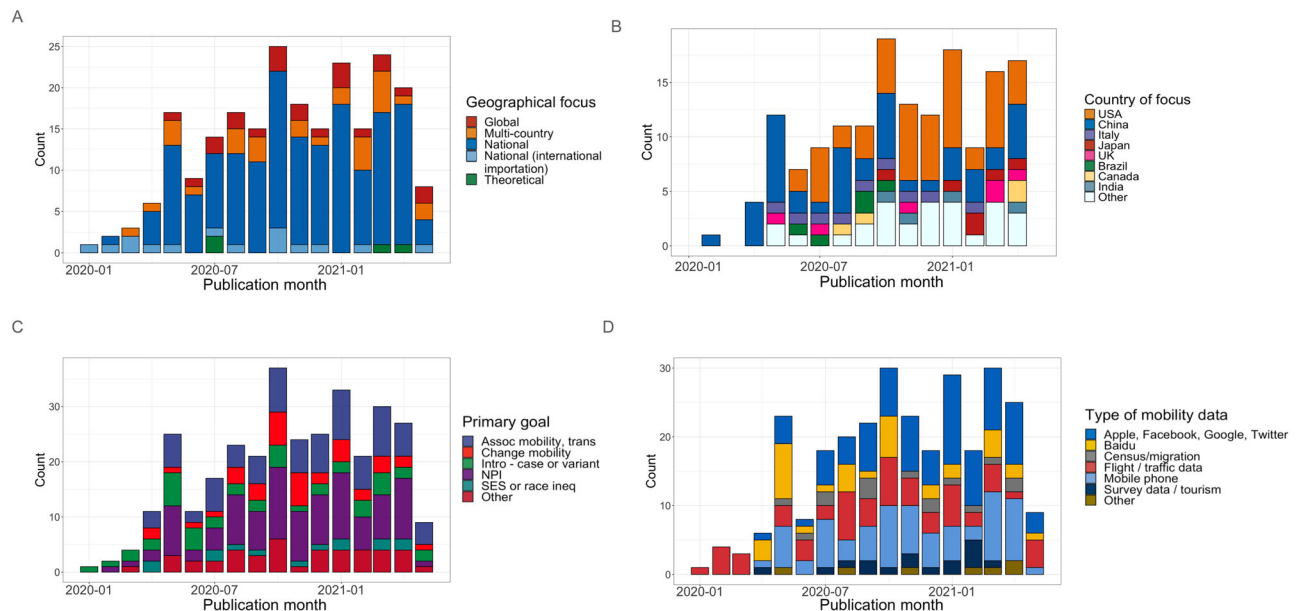


Fig. 1 | Geographical and country focus, primary goal, and type of mobility data used in papers published between January 2020 and May 9, 2021, included in the systematic review (n=232). **A** Geographical focus for papers published between January 2020 and May 9, 2021. National focus refers to analysis focused on one country at national or subnational level. **B** Distribution of countries of focus for papers published between January 2020 and May 9, 2021, for the eight countries

with the most papers included in the analysis. **C** Primary goal of the study. Assoc mobility, trans: Quantifying association between mobility and transmission of SARS-CoV-2; Change mobility: Change in mobility during pandemic; Intro—case or variant: Introduction or importation of SARS-CoV-2 cases or variants; NPI: Effectiveness of non-pharmaceutical interventions; SES or race ineq: Socio-economic or racial inequities. **D** Types of mobility data used, by date.

Table 1 | Results of statistical testing for change of variable of interest over time

Test	p-value	Categories of outcome variables included
Geographic focus vs time	0.087	Import, National, Multi-country, Global, Theory
Primary goal vs time	0.185	Intro—case or variant, NPI, Other, Assoc mobility, trans, Change mobility, SES or racial inequality
Mechanistic/statistical integration vs time	0.607	Mechanistically, Statistically
Type of mobility data vs time	0.185	Flight/traffic data, Apple/Facebook, Google/Twitter, Baidu, Mobile phone, Other

For each outcome of interest with more than two levels (geographic focus, primary goal, and type of mobility data), we fit a multinomial regression, with time as independent variable, and carried out Anova test to obtain the overall p-value. For the binary outcome (mechanistic/statistic integration vs time), we fit a logistic regression with time as independent variable.

case introductions and one on variants. Only 14 (6%) papers explored associations between socio-economic status (SES) or racial inequalities and mobility. There was not enough evidence to conclude that the overall goal of the paper changed over time (Fig. 1C, Table 1). Primary goals were generally similar in papers that used single vs. multiple datasets for mobility, although papers with more than one dataset more commonly focused on SES and racial inequities (7, 17%) (see Supplementary Table 3).

Section 2: Mobility data analysis

Mobility data. The majority (191, 82%) of papers analyzed a single mobility data set, and no paper analyzed more than four mobility datasets. The Google Community Mobility Reports (52 occurrences) was the most used type of data, followed by airline flight data (40 occurrences) and Baidu data (35 occurrences). Multiple types of mobile phone data were collected via applications (19 occurrences), roaming (2 occurrences) or GPS locator/Call detail records (36 occurrences). In some instances, particular type of mobile phone data was not specified (6 occurrences). Mobile phone data collected via applications included data collected from applications like Tencent, Wayz Inc., Cuebiq Inc., CityMapper, PlacelQ, and SafeGraph. While data from social media platforms like Twitter/X and Facebook were likely most commonly collected via mobile phone applications, we considered them to be more similar to other types of digitally

collected data, namely Apple and Google data. In 15 occurrences, authors used migration or census data; migration data were defined as data on migration flows between administrative units, usually defined by governments¹⁹, while census data captured mobility based on change of residence from 1 year prior to time of the census²⁰.

When combining mobile phone data collected via applications, roaming, GPS locator/Call data records, and other types of mobile phone data (excluding Twitter/X) into larger, less granular categories, mobile phone data were the most used dataset (72 occurrences). Less commonly used mobility data, with at least two occurrences for each, included traffic data, include census/migration/commuting data, Apple application, Facebook, survey, Google App, and Twitter data, travel histories, tourism logs, and a composite data set provided by the University of Maryland.

Of the 42 papers that used two or more datasets, there were no consistent patterns in the combinations of data types used. There were no clear trends in the type of mobility data used over time (Table 1), although the papers published during the first three months of the pandemic (versus the entire time period analyzed) used only flight/traffic data (Fig. 1D). There was evidence of an association between the type of mobility data used and the geographic area of focus ($p < 0.001$) (Table 2). Notably, while national focus was the most common focus for all types of mobility data, there was more variation in focus for Apple, Facebook, Google, and Twitter data compared to the other

Table 2 | Results of statistical testing using Anova test for association between variables

Test	p-value	Variable 1 categories	Variable 2 categories
Type of mobility data vs geographic focus of paper	<0.001	Type of mobility data: Apple, Facebook, Google, Twitter; Baidu; Flight/traffic data; Mobile phone data; Other	Type of geographic focus of paper: Global; Multi-country; National; National (international importation) Note: Theoretical focus excluded due to <5 instances
Use of statistical/mechanistic model vs primary goal of paper	0.0537	Model used in integration: Statistical; Mechanistic	Primary goal of paper: Quantifying association between mobility and transmission of SARS-CoV-2; Effectiveness of non-pharmaceutical interventions; Change in mobility during pandemic; Introduction or importation of SARS-CoV-2 cases/variants; Socio-economic or racial inequities; Other

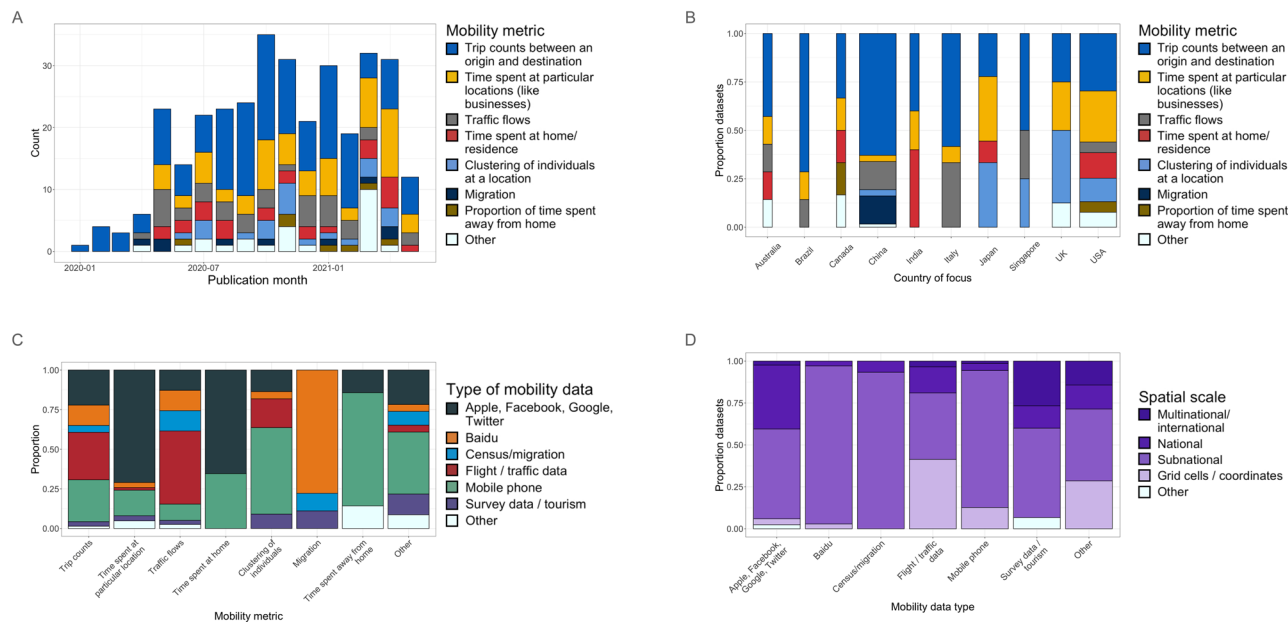


Fig. 2 | Mobility metrics, type of mobility data, and spatial scale used in mobility datasets in the papers analyzed ($n = 232$). **A** Distribution of mobility metrics used, by date of publication. **B** Mobility metric by country of paper focus, for the ten most commonly used countries. Width of the bars represents the relative number of mobility metrics presented for that country (Australia, Brazil,

Canada, India, and Singapore with less than 8 metrics; Italy, Japan, and the UK with 8–12 metrics; and China and the USA with 62 and 91 metrics, respectively). **C** Type of mobility data, by mobility metric. **D** Spatial scale by mobility data type. Subnational scale includes city/village/metro area, province/state, county/district, and smaller than city/village/metro area.

datasets. The proportion of flight/traffic data that focused on international importation was the highest compared to the other data types. None of the papers with a global focus used mobile phone data.

Mobility metrics. Various measures of mobility (mobility metrics) were analyzed. The most common mobility metric used was the calculated trip count between an origin and destination (141 occurrences, 43%), followed by time spent at particular locations (63 occurrences, 19%), and traffic flows (39 occurrences, 12%) (see Supplementary Table 4). Less common mobility metrics included time spent at home or away, clustering of individuals at a location, migration and other roaming, migration and international trips, and tourism statistics (see Supplementary Table 4). While in the first three months, only trip counts between origin and destination were used, over time, there was more variability in mobility metrics used, with trip counts between origin and destination making up a smaller proportion of metrics as the pandemic progressed (Fig. 2A). When looking at analyses in the two most commonly researched countries (USA and China), a wider range

of metrics were used in USA-based analyses than in China, with almost 60% of China-focused papers using trip counts (Fig. 2B). This may be related to the predominance of a single source of mobility data analyzed from China, with most data coming from Baidu and flight/traffic data. Metrics were not equally calculated across all data types. For example, certain metrics were predominantly available from particular datasets, such as migration information from Baidu or time spent away from home from mobile phone data (Fig. 2C). Apple, Facebook, Google, and Twitter data were used to calculate the largest number of different metrics, and were most commonly used to calculate time spent at particular locations, and trip counts between origin and destination (Table 3).

Spatial scale. Mobility data available at subnational level, specifically city, village, or metro area, were the most commonly used type of data (59 occurrences, 20%), followed by province/state-level (e.g., administrative level 2 data) (52 occurrences, 18%), national-level data (e.g., administrative level 1 data) (49 occurrences, 17%), and other types of

Table 3 | Mobility data characteristics, by type of data set

Characteristic	Apple, Facebook, Google, Twitter, N = 77	Baidu, N = 33	Census/migration, N = 13	Flight/traffic data, N = 5	Mobile phone data, N = 62	Survey data/tourism, N = 16	Other, N = 6
Mobility metric							
Clustering of individuals at a location	2.8%	2.9%	0%	6.1%	14%	15%	0%
Migration	0%	21%	7.1%	0%	0%	7.7%	0%
Other	4.7%	2.9%	14%	1.5%	10%	23%	22%
Proportion of time spent away from home	0.9%	0%	0%	0%	5.8%	0%	11%
Time spent at home/residence	16%	0%	0%	0%	10%	0%	0%
Time spent at particular locations (like businesses)	42%	5.9%	0%	1.5%	12%	15%	33%
Traffic flows	4.7%	15%	36%	27%	4.7%	7.7%	11%
Trip counts between an origin and destination	29%	53%	43%	64%	43%	31%	22%
Unknown	2	3	2	1	6	3	2
Spatial scale							
Multinational/international	2.4%	0%	0%	3.4%	1.4%	27%	14%
National	38%	2.9%	6.7%	16%	4.2%	13%	14%
Subnational	54%	94%	93%	40%	82%	53%	43%
Grid cells/coordinates	3.6%	2.9%	0%	41%	13%	0%	29%
Other	2.4%	0%	0%	0%	0%	6.7%	0%
Unknown	0	0	0	0	1	1	0
Granularity							
Data aggregates multiple individuals' movements	90%	100%	100%	95%	92%	40%	71%
Measures individual-level movements	9.5%	0%	0%	5.2%	8.5%	60%	29%
Unknown	0	0	0	0	1	1	0
Details on population available	12%	2.9%	13%	12%	11%	47%	0%
Unknown	0	0	0	0	1	1	0
Direct metadata available							
None	99%	100%	65%	90%	89%	36%	100%
Age	0%	0%	5.9%	6.9%	6.8%	20%	0%
Ethnicity/race	0%	0%	5.9%	1.7%	0%	4.0%	0%
Gender/sex	0%	0%	18%	1.7%	1.4%	16%	0%
Socio-economic Information	0%	0%	0%	0%	2.7%	20%	0%
Other—clinical	1.2%	0%	0%	0%	0%	4.0%	0%
Other—Means of transport, duration of journey, reason for moving	0%	0%	5.9%	0%	0%	0%	0%
Ecological metadata available							
Age	1.1%	0%	24%	3.3%	7.1%	6.7%	0%
Ethnicity/race	2.3%	0%	9.5%	1.7%	6.0%	0%	0%
Gender/sex	1.1%	0%	14%	1.7%	4.8%	0%	0%
None	92%	97%	43%	93%	70%	87%	100%
Socio-economic information	3.4%	2.9%	9.5%	0%	12%	6.7%	0%
Information about bias from sampling							
No, the authors do not provide this information	75%	86%	87%	83%	74%	73%	86%
Not applicable	1.2%	2.9%	6.7%	1.7%	0%	0%	14%
Yes, the authors provide information and use it to adjust estimates of mobility	7.1%	0%	0%	8.6%	10%	0%	0%
Yes, the authors provide information but do not use it to adjust estimates of mobility	17%	11%	6.7%	6.9%	16%	27%	0%

Percentages presented are by type of data set used (that is, column-wise proportions).

Mobile phone data are inclusive of data from mobile phone applications, billing records, and GPS locations.

subnational level data (45 occurrences of county/district level data, and 30 occurrences of data at smaller than city/village/metro area level) (see Supplementary Table 5). Less commonly, the spatial scale was airport-to-airport, longitude/latitude coordinates, multinational/international, or grid cell level. Two spatial scale levels were used in only two papers (<1%). While airport-to-airport data were the earliest used spatial scale of analysis, subnational and national scale data were used quickly after (within three months). It took six months before data at a multinational scale were used (Supplementary Fig. 5). Some types of mobility data were dominated by a specific spatial scale; for example, Baidu data were almost always at the subnational level, whereas flight/traffic data were more commonly analyzed at the grid cell/coordinates level (Fig. 2D).

Granularity of data. The vast majority of the time, mobility data used were an aggregate of multiple individuals' movements (259 out of 288 occurrences, 90%). Only aggregated data were used in the first three months of 2020 (Supplementary Fig. 6A). Unlike the rest of data types, most datasets using survey data/tourism measured individual-level movements (Supplementary Fig. 6B). Most datasets focused on daily mobility patterns; sub-daily mobility was primarily quantified using Facebook and other app data and to a lesser extent mobile phone data and traffic data (Supplementary Table 20, Supplementary Fig. 7). Higher levels of temporal aggregation, often monthly or longer, were more common when using flight/traffic data, survey data, and census/migration data (Supplementary Table 20, Supplementary Fig. 7).

Details on population from which data were collected. Very few papers provided any information on the population from which mobility data were collected, including age, gender/sex, race/ethnicity, or other socio-economic information (35 out of 288 mobility datasets, 12%). This was consistent across the entire study period. Survey data were the only data type where population information was provided the majority of the time (Table 3, Supplementary Fig. 8A); metadata were especially rare for mobile phone data, Baidu, and data from Apple, Facebook, Google, and Twitter, although with notable exceptions²¹. As a result, very few (21, 7%) datasets had metadata available directly for participants for which mobility data were collected (at individual scales); among those that did, age was the most common type of metadata available (15 occurrences), followed by gender/sex (9 occurrences), socio-economic information (7 occurrences), and ethnicity/race (3 occurrences). Clinical metadata were available in 2 occurrences, and metadata on travel (means of transport, duration of journey, and reason for moving) was available in 1 occurrence (Supplementary Table 6). Notably, socio-economic information was only available in datasets collected through mobile phone or survey, while gender/sex information was also available from census/migration and traffic data (Table 3, Supplementary Fig. 8B). Metadata were available ecologically (i.e., for the population at large for which mobility applies) at a similarly infrequent rate; when available, socio-economic information was most common (17 occurrences, 6%), followed by age (15 occurrences, 5%), ethnicity/race (10 occurrences, 3%), and gender/sex (9 occurrences, 3%) (Supplementary Table 7). Census/migration data had the largest proportion of reporting metadata ecologically (Table 3, Supplementary Fig. 8C).

In most instances, authors did not provide information about possible bias in sampling or in the data collection mechanism (226 occurrences, 79%). When the authors acknowledged potential bias, it was usually limited to a mention or description of the population that was included or excluded from the sample, without using this information to adjust estimates of mobility (38 occurrences, 13%). The authors provided information on the sample and used it to adjust estimates in 18 occurrences (6%) (Supplementary Table 8). Adjustment was reported when mobility data were airline/flight data, Apple application data, Facebook, Google, mobile phone, and traffic data

(Table 3, Supplementary Fig. 8D). In 90% of papers (209 out of 231 papers), authors did not report internal or external validation of mobility data, with only 14 papers (6%) reporting that the data were validated for external consistency with another data source, and 8 papers (3%) reporting that data were validated for internal consistency (Supplementary Table 22). Finally, mobility patterns generally could not be matched to individuals' disease status. Matching was only possible for 13 out of 288 occurrences (5%). Linking mobility to disease status was most common in survey and tourism data (Supplementary Fig. 9).

Availability of mobility data. Most often, authors used mobility data that were available for free from an open source (136 occurrences, 47%). This was followed by data made available to authors by a company, government agency, and other entities (94 occurrences, 33%). In terms of secondary uses of the data following the authors' publication, less commonly, data were available by request to the authors (8 occurrences, 3%) or were available for purchase (5 occurrences, 2%) (Supplementary Table 9). For 42 occurrences, authors did not provide information on data availability (15%). Interestingly, data available for purchase were not used in the first seven months considered, during which published papers all used open-source data and data made available by a company, government agency, and other entities (Supplementary Fig. 10).

Identifiability or privacy of data. While in many occurrences, information about identifiability or privacy of data was not explicitly stated (61 occurrences, 21%), most commonly, mobility data were anonymized and spatially and temporally aggregated (191 occurrences, 67%). Data were anonymized and spatially aggregated in 21 occurrences (7%), followed by the removal of identifiers (9 occurrences, 3%) and anonymization and temporal aggregation (2 occurrences, 1%) (Supplementary Table 10). There appears to be no temporal variation in the identifiability or privacy of the data used (Supplementary Fig. 11).

Section 3: COVID-19/SARS-CoV-2 data analysis

Disease data are key to understanding the impact of human mobility data on transmission since they provide a path forward to validate the use of mobility data. Given the abundance and availability of SARS-CoV-2 data, we further explored the diversity and types of data analyzed in conjunction with mobility data.

Type of SARS-CoV-2 data used. Most papers used confirmed cases (178 occurrences, 77% of papers) of COVID-19, followed by confirmed deaths (59 occurrences, 26%). Less commonly, SARS-CoV-2 data were drawn from suspected cases (9 occurrences, 4%), hospitalizations (9 occurrences, 4%), suspected deaths (6 occurrences, 3%), serological or virological results (4 occurrences, 2%), or other (22 occurrences, 10%) (Supplementary Table 11). Only 28 papers (12%) did not use SARS-CoV-2 data. In the first three months, papers relied on confirmed cases only, or did not use SARS-CoV-2 data (Supplementary Fig. 12).

By far, the most common source of SARS-CoV-2/COVID data was government or public health reports (124 occurrences, 61%). The JHU COVID-19 Dashboard/JHU Coronavirus Resource Center was used in 35 occurrences (17%), followed by non-governmental organization report or repositories (21 occurrences, 10%), media including repositories (19 occurrences, 9%), and WHO reports (18 occurrences, 9%). Less commonly, research studies, including active and passive data collection, hospital and health facility records, other university repositories, and previous studies were used as sources of these data (Supplementary Table 12). There was some variability in the source of data by country of study focus; for the top ten countries of focus, government and public health reports were the predominant sources of SARS-CoV-2/COVID-19 data, except the USA, which had more variability in the source of data (Supplementary Fig. 13).

In most cases, there was a complete spatial overlap between the mobility data and SARS-CoV-2 data (141 papers, 71% of papers reporting response to this question) (Supplementary Table 13). Temporally, there was often a mismatch between the datasets, with only 37% of data having complete temporal overlap between mobility and SARS-CoV-2 data (75 papers) and 16% having no overlap (33 papers) (Supplementary Table 14). Of papers with no temporal overlap ($n = 33$), these papers most often used flight/traffic data (16 occurrences, 42%) or census/migration data (12 occurrences, 32%) (Supplementary Table 15).

Section 4: Integration of mobility data and SARS-CoV-2 data/transmission

Of the 143 papers that modeled SARS-CoV-2 transmission, most commonly, a statistical model was used (81 occurrences, 57%), followed by a mechanistic model (64 occurrences, 45%). Two papers used both statistical and mechanistic approaches in their analysis. When statistical models were used to link mobility and SARS-CoV-2 data, they often used a simple regression model (55 papers) or calculated a correlation metric (41 papers). Less commonly, they used network models (9 papers), summary statistics (9 papers), or non-linear regression (5 papers). Of the 43 papers that used a mechanistic model, SEIR/SIR type models including spatial structure were the most common (20 papers), followed by models with contact (9 papers) structure, or other approaches (Supplementary Table 16, Fig. 3A). There was no relationship between the primary question (NPIs, change in mobility, etc.) and the use of either statistical or mechanistic models ($p = 0.0537$) (Table 2). Over time, there was a small relative decline in the use of mechanistic models compared to statistical models in integrating mobility and SARS-CoV-2; however, the change was not statistically significant (Table 1).

When using mobility to statistically or mechanistically predict changes in SARS-CoV-2/COVID-19 infection, cases, hospitalizations, and/or deaths, papers usually provided measures of prediction accuracy but very rarely evaluated accuracy with and without mobility data (Supplementary Fig. 14). Evaluating accuracy was somewhat more common in the papers where mechanistic prediction using mobility data was done (Fig. S18A, B); however, the number of papers where accuracy was evaluated was very small (2 with statistical prediction and 5 with mechanistic prediction).

Models of mobility. Few papers explicitly performed any modeling of the mobility data directly. We only found six instances where common spatial interaction models were used to model trends in mobility (5 = gravity, 1 = radiation). Statistical models of mobility were more common (51 times, 22%), although most papers did not model mobility (67%) (Supplementary Table 21). Papers rarely evaluated a null hypothesis, i.e., the mobility data were needed for their analysis. In total, 80% of papers (184 papers) did not evaluate a null hypothesis.

Section 5: Reported results

Given the wide diversity of topic areas, analysis types, and modeling approaches, there was no consensus in reported results across papers. We focused on results explicitly providing a quantitative understanding of how changes in mobility were impacted following the implementation of lockdowns (stay-at-home orders) and travel restrictions, as well as how change in mobility was associated with change in SARS-CoV-2 transmission.

When looking at papers focused on evaluating the effectiveness of lockdown (causally attributing change in cases, deaths, R0, etc. to lockdown) (58 papers), most commonly, the results presented included the effectiveness of NPIs on SARS-CoV-2/COVID-19 (79%). Less commonly, these papers presented forecasts or projections of SARS-CoV-2/COVID-19 infections, cases, hospitalizations, and/or deaths (34%), followed by estimates of basic reproductive number (28%), and

prediction/comparison of NPI strategies (26%) (Supplementary Table 17). In most cases, papers found a decrease in the number of cases, deaths, or R0 following lockdown measures (Fig. 3B), followed by mixed associations, where the direction of association between lockdown and cases, deaths, and basic reproductive number varied across time periods, space, or groups of interest (Fig. 3B). Decreases in COVID-19 burden were observed regardless of the type of mobility data analyzed for all types of data that had at least five occurrences (i.e., excluding survey data/tourism, which only had one occurrence) (Supplementary Fig. 15). For papers that reported an association (not necessarily a causal relationship) between mobility and lockdown measures (71 papers), most reported decreased mobility after lockdown, followed by decreased mobility before lockdown and mixed associations (Supplementary Fig. 16). These results did not appear to change over time (Supplementary Fig. 16) and were largely consistent by type of mobility data; however, given the low number of occurrences by type of mobility data that looked at the association between mobility and lockdown, results should be interpreted with caution.

Among papers that assessed the effectiveness of travel restrictions (50 papers), most analyses reported measures of the effectiveness of NPIs on COVID-19 (69%), followed by forecasts or projections of SARS-CoV-2/COVID-19 infections, cases, hospitalizations, and/or deaths (35%), prediction/comparison on NPIs (25%), and basic reproductive number (17%) (Supplementary Table 18). Like analyses focused on lockdown measures, the implementation of travel restrictions was generally associated with decreased cases, deaths, or R0 (Fig. 3C), followed by mixed associations. This did not appear to change over time (Fig. 3C). Decreased cases, deaths, and basic reproductive number following travel restrictions were the most common results observed no matter the type of mobility data used (Supplementary Fig. 17).

More than half of the papers (127 papers) evaluated whether there was an association between the change in mobility and transmission of SARS-CoV-2/COVID-19. Most commonly, papers reported an association between change in mobility and disease transmission (59% reported decreased mobility and decreased transmission; 28% reported increased mobility and increased transmission), followed by mixed associations (20% of papers) (Supplementary Fig. 18, Supplementary Table 19). This did not appear to vary over time (Fig. 3D).

As an example of the diversity of approaches and types of analyses investigating mobility data and SARS-CoV-2 transmission, we analyzed the national papers that explicitly reported or used effective reproductive number (R_t) in their analysis ($n = 30$). Among these, 9 papers focused explicitly on comparing mobility to an estimate of R_t , the time-varying R_0 value, in the USA. Of these papers, two focused on superspreading events, with the remaining focused on the relationship between R_t and mobility, and implementation of NPIs such as lockdowns. While the majority of these papers focused on subnational mobility data analyzed on a daily resolution, they greatly differed in their approaches. For example, half linked SARS-CoV-2 transmission and mobility statistically ($n = 16$), and the remaining papers focused on a mechanistic approach using a range of formulations (including age: 4, contact: 2, spatial: 3, individual-based: 4, semi-mechanistically: 2). Results presented in the nine papers focused on the US did not consistently report how changes in mobility or the amount of mobility would translate to changes in R_t . For example, Unwin et al.²², Parker and Pianyk²³, and Liu et al.²⁴ all report relationships between mobility and R_t , but these approaches all differ in how R_t was calculated, how mobility was defined, the methods for relating R_t and mobility, and the subsequent translation of these results.

Discussion

Mobility data can help us to better understand the relationship between human behavior and disease transmission. During the COVID-19 pandemic, there was a flurry of analyses and increased data

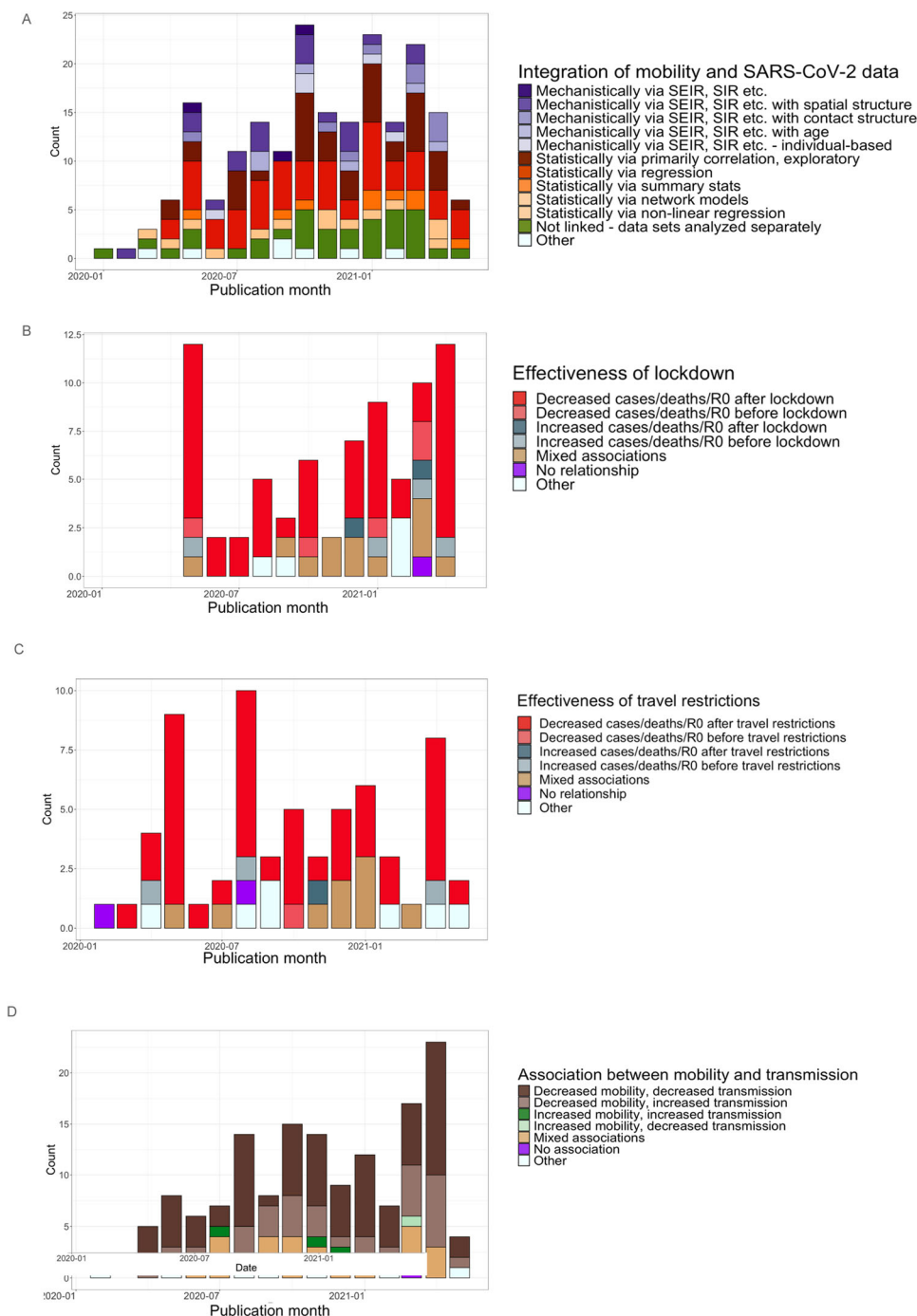


Fig. 3 | Ways in which mobility data and data on SARS-CoV-2 were integrated, and results presented in the papers ($n = 232$). **A** Integration of mobility data and data on SARS-CoV-2/COVID-19, by date. **B** Results presented in papers assessing effectiveness of lockdown measures. Lockdown includes lockdown, social distancing, and stay-at-home orders. Mixed associations refer to associations that vary

across periods, space, or groups of interest. **C** Results presented in papers assessing effectiveness of travel restriction measures. Mixed associations refer to associations that vary across periods, space, or groups of interest. **D** Results presented in papers assessing relationship between change in mobility and transmission of SARS-CoV-2/COVID-19.

availability to describe the mobility patterns of individuals and populations globally. We conducted a systematic review of these papers and found that multiple types of data were used across geographic settings for a range of different analytical purposes. In general, mobility data were analyzed in conjunction with SARS-CoV-2/COVID-19 data to explore the effectiveness of NPIs, to quantify changes in mobility during the pandemic, and to understand the relationship between mobility and disease burden. The methods by which mobility data were integrated with SARS-CoV-2/COVID-19 data/models or analyzed

independently were not observed to be consistent across articles included in our review.

While many of these papers were focused on using information for public health policy, there were few papers with a co-author affiliated with a public health agency. While there may be many causes for this, including authorship restrictions, challenges around integration between academia and public health officials, and scientific publications being only one mechanism for collaboration, this result could indicate an important gap between research and translation of findings

into public health practice. In contexts such as the COVID-19 pandemic, closer collaboration between academia and policymakers could help bridge these gaps.

Data extraction, especially details about mobility data, was challenging due to the lack of standardized reporting across journals and papers, underscored by the high number of journals in which the papers were published. Authors may not have explicitly listed particular pieces of information, such as duration of analysis, information about the population whose mobility was described by the data, or spatial and temporal scale of data used. The lack of consistent reporting of analysis approaches and how mobility data were integrated (or not) with estimates of SARS-CoV-2 transmission contributed to the difficulty in identifying consistent patterns and relationships between mobility and SARS-CoV-2 transmission. Establishing guidelines for reporting pre-processing and processing of mobility data, their use in analysis, and the way results are presented could mitigate some of these issues and facilitate reproducibility and communication of findings²⁵.

Consistent reporting of mobility data, including information about the population from which data were collected, could also aid in mitigating potential sources of bias. Most papers used data from applications originally developed for purposes other than collecting mobility data, such as Facebook data, Google, Twitter, Apple, or mobile phone data. Open questions remain about who is captured in these datasets and how these biases may systematically exclude particular populations. For example, while mobile phone ownership is generally increasing, phone ownership may vary by age, sex, socioeconomic demographics, and other factors^{26–34}. There is also evidence suggesting demographic and behavioral variation in population cohorts who use social media³⁵. While mobility data were often freely and publicly available during the pandemic, they rarely included any metadata. As a result, few papers (6%) adjusted for any biases in sampling before analysis or carried out any validation of mobility data. The extent to which potential biases in mobility data could affect the findings of the studies remains to be seen, as few studies explicitly tried to adjust for these factors.

This study has several limitations. First, our search strategy and analyses included papers published only between the beginning of the pandemic and May 2021. This time frame was used to restrict our analysis to papers published when our understanding of mobility and SARS-CoV-2 transmission was rapidly developing. At the beginning of this period, there was very little known about SARS-CoV-2 transmission in comparison to the period when clear relationships were assumed between these factors. In addition, this was the period when rapidly conducted analyses were being used to guide public health practice and policy before widespread vaccination became available. Finally, these papers represent the time when the science and approaches were being newly developed, and we wanted to explore how this would change as our understanding of mobility and SARS-CoV-2 changed, during the specified period. We anticipate that following this period, the availability of mobility data would increase, and there would likely be an increase in the diversity of countries that were the focus of papers that included mobility data. We would expect that after this period, the primary goal of research papers would have shifted towards the effectiveness of NPIs and understanding changes in risk associated with emergent SARS-CoV-2 variants. We also expect that papers published after our study period were more likely to use large-scale surveys, census, and other types of mobility data that require a significant time commitment for collection, processing, and making data available. We excluded papers that used mobility data on less than 1000 individuals; we had used this threshold to delineate papers with population-level mobility from case studies and individual case investigations. However, this could have resulted in the exclusion of some smaller surveys, although only 15 papers were

excluded based on this criterion. In addition, we restricted our search to publications published in English. This may have resulted in a biased picture of all research conducted since COVID-19 was a global concern. In addition, nationally focused papers were dominated by two countries, the USA and China, with few coming from the global South.

This systematic review represents the vast amount of research on the relationship between mobility and SARS-CoV-2 transmission. These papers analyzed large quantities of mobility data, much of it publicly available, which has been unprecedented on a global landscape. However, to leverage these findings into both a better scientific understanding and actionable public health policy, there are a number of steps that could be implemented. For example, having standardized, systematic ways to use of mobility data in analyses—from selecting a dataset to thinking about and reporting potential biases, to using appropriate methods to integrate these datasets, and conducting statistically rigorous evaluations of these results—could improve the overall reproducibility, interpretability, and reusability of results beyond SARS-CoV-2.

Methods

Paper screening, review, and data extraction

We screened abstracts published from 2019 to the date of search (May 9, 2021) across publication databases Pubmed and EMBASE, using keywords and MESH terms about SARS-CoV-2/COVID-19 and travel/mobility (see Supplementary Methods 1 for the complete search list). Keywords that focused on travel/mobility also included modes of transport (e.g., cruise, airline, train, etc.) and general words about travel (e.g., travel, mobility, movement). Preprints and peer-reviewed literature were included if they were available in one of the repositories searched. However, in the case that both were available, the peer-reviewed version was included.

Each abstract was screened independently by two researchers and categorized into one of six categories: papers that included data on mobility to characterize SARS-CoV-2 transmission; papers about SARS-CoV-2 that did not focus on mobility or were case reports; papers not primarily about SARS-CoV-2; no abstract or erratum/correction; bad return; different language; and not original research articles. A description of each category is available in Supplementary Methods 2. Cases of disputes in categorization were resolved by agreement from two independent researchers.

Abstracts that were categorized as using quantitative data on population-level human mobility to characterize SARS-CoV-2 transmission (first category of abstracts) were eligible for full review and data extraction using an online survey software (Survey Monkey)³⁶. Some additional papers were excluded from analysis if, during the full review, they were identified as not having used mobility data, having used information on mobility from a small population (less than 1000 people), or having been published outside the period of interest. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram is available in Supplementary Fig. 1.

Full reviews and data extraction were carried out on papers included for systematic review by twenty-four independent researchers. Broadly, data items collected included general data about the study, including information about the authors of the papers and their affiliation, primary objectives of the study, geographical focus, spatial scale, and time frame of analysis. We collected information about mobility data used in the study, including types of mobility data, source, spatial and temporal scale, mobility metrics, and data availability. We extracted information about SARS-CoV-2/COVID-19 data used in the paper, how disease and mobility data were integrated, and what results were presented. The complete version of the survey questionnaire is available in the Supplementary Methods 3.

While we did not formally assess the risk of bias for the papers included in the analysis, for each of the papers, we extracted data

BOX 1

Overview of domains of interest for the analysis of papers using population-level human mobility data to understand SARS-CoV-2 transmission

Attributes on the focus/goal of the paper, mobility data, type of analysis, use for understanding SARS-CoV-2 transmission and reported results were extracted from each paper using a standardized survey with the following sections.

- Section 0: Basic metadata on published papers—date of publication, journal, and information on authorship.
- Section 1: Area of focus—geographic area of interest, temporal and spatial scale of data and analysis, and the primary goal of the paper.
- Section 2: Mobility data analysis—data types analyzed, availability of and access to datasets; spatial, temporal, and population scale of mobility data; and the method of analysis

- Section 3: COVID-19/SARS-CoV-2 data analysis—COVID-19/SARS-CoV-2 data types analyzed; source of data; spatial and temporal overlap between COVID-19/SARS-CoV-2 data and mobility data
- Section 4: Integration of mobility data and SARS-CoV-2 data/transmission— modeling approach; method to integrate mobility data and SARS-CoV-2 data/analyses
- Section 5: Reported results—type of results reported in the papers, metrics of mobility, SARS-CoV-2 analyses

on the availability of individual and ecological metadata on the population that was included in the mobility data and whether authors used this information to correct for potential sampling bias.

Data analysis

For most data points of interest, we calculated descriptive statistics like absolute and relative frequencies for categorical variables (e.g., types of mobility metrics reported), or mean, median and interquartile ranges for continuous variables (e.g., duration of study analysis). To test change in categorical variables over time (e.g., primary goal of papers reviewed), we fit a multinomial regression, with date of publication as independent variable. We used the Anova test to obtain an overall p -value. For binary outcome variables (e.g., statistical vs mechanistic integration), we fit a logistic regression, with date of publication as independent variable. The association was considered statistically significant if the p -value was <0.05 .

An overview of domains of interest for the analysis is provided in Box 1.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The dataset generated and used in the analysis is available for download on GitHub (https://github.com/nkostandova/covid_mobility_sysreview).

Code availability

All code used for cleaning, generation of results, and visualization is provided in the same repository as data (https://github.com/nkostandova/covid_mobility_sysreview).

References

- Chen, W. et al. [Early containment strategies and core measures for prevention and control of novel coronavirus pneumonia in China]. *Zhonghua Yu Fang Yi Xue Za Zhi* **54**, 239–244 (2020).
- Garcia, L. P. & Duarte, E. Nonpharmaceutical interventions for tackling the COVID-19 epidemic in Brazil. *Epidemiol. Serv. Saude* **29**, e2020222 (2020).
- Heymann, D. L. & Shindo, N. COVID-19: what is next for public health? *Lancet* **395**, 542–545 (2020).
- Lee, V. J., Chiew, C. J. & Khong, W. X. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. *J. Travel Med.* **27**, taaa039 (2020).
- Zheng, Q. et al. HIT-COVID, a global database tracking public health interventions to COVID-19. *Sci. Data* **7**, 286 (2020).
- Cheng, C., Barceló, J., Hartnett, A. S., Kubinec, R. & Messerschmidt, L. COVID-19 government response event dataset (CoronaNet v.1.0). *Nat. Hum. Behav.* **4**, 756–768 (2020).
- Etemad, K. et al. Non-pharmacologic interventions in COVID-19 pandemic management; a systematic review. *Arch. Acad. Emerg. Med.* **11**, e52 (2023).
- Barbosa, H. et al. Human mobility: models and applications. *Phys. Rep.* **734**, 1–74 (2018).
- Hu, T. et al. Human mobility data in the COVID-19 pandemic: characteristics, applications, and challenges. *Int. J. Digit. Earth* **14**, 1126–1147 (2021).
- Huang, X. et al. The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the U.S. during the COVID-19 pandemic. *Int. J. Digit. Earth* **14**, 424–442 (2021).
- Huang, X. et al. Time-series clustering for home dwell time during COVID-19: what can we learn from it? *ISPRS Int. J. Geo-Inf.* **9**, 675 (2020).
- Google. *COVID-19 Community Mobility Report*. <https://www.google.com/covid19/mobility?hl=en> (2024).
- Huang, X., Li, Z., Jiang, Y., Li, X. & Porter, D. Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PLoS ONE* **15**, e0241957 (2020).
- Warren, M. S. & Skillman, S. W. Mobility changes in response to COVID-19. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.2003.14228> (2020).
- Grantz, K. H. et al. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat. Commun.* **11**, 4961 (2020).
- Ilin, C. et al. Public mobility data enables COVID-19 forecasting and management at local and global scales. *Sci. Rep.* **11**, 13531 (2021).
- Zhang, M. et al. Human mobility and COVID-19 transmission: a systematic review and future directions. *Ann. GIS* **28**, 501–514 (2022).
- World Bank. *World Bank Country and Lending Groups—World Bank Data Help Desk*. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups> (2024).

19. Zhan, C., Tse, C. K., Lai, Z., Chen, X. & Mo, M. General model for COVID-19 spreading with consideration of intercity migration, insufficient testing, and active intervention: modeling study of pandemic progression in Japan and the United States. *JMIR Public Health Surveill.* **6**, e18880 (2020).
20. United States Census Bureau. *American Community Survey and Puerto Rico Community Survey: 2022 Subject Definitions*. https://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/2022_ACSSubjectDefinitions.pdf (2022).
21. Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S. & Colizza, V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. *Lancet Digit. Health* **2**, e638–e649 (2020).
22. Unwin, H. J. T. et al. State-level tracking of COVID-19 in the United States. *Nat. Commun.* **11**, 6189 (2020).
23. Parker, D. & Panykh, O. Mobility-guided estimation of COVID-19 transmission rates. *Am. J. Epidemiol.* **190**, 1081–1087 (2021).
24. Liu, H., Chen, C., Cruz-Cano, R., Guida, J. L. & Lee, M. Public compliance with social distancing measures and SARS-CoV-2 spread: a quantitative analysis of 5 states. *Public Health Rep.* **136**, 475–482 (2021).
25. Kostandova, N. et al. Data and model needs for generalizable inferences linking human mobility and infectious disease transmission. Preprint at OSF <https://doi.org/10.31219/osf.io/y9mzd> (2024).
26. Mao, L., Yin, L., Song, X. & Mei, S. Mapping intra-urban transmission risk of dengue fever with big hourly cellphone data. *Acta Trop.* **162**, 188–195 (2016).
27. Frias-Martinez, V., Soguero, C. & Frias-Martinez, E. Estimation of urban commuting patterns using cellphone network data. In *Proc. ACM SIGKDD International Workshop on Urban Computing*, 9–16 (Association for Computing Machinery, 2012).
28. Silver, L. & Johnson, C. *Internet Connectivity Seen as Having Positive Impact on Life in Sub-Saharan Africa: But Digital Divides Persist*. <https://www.pewresearch.org/global/2018/10/09/majorities-in-sub-saharan-africa-own-mobile-phones-but-smartphone-adoption-is-modest/> (Pew Research Center, 2018).
29. Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W. & Buckee, C. O. The impact of biases in mobile phone ownership on estimates of human mobility. *J. R. Soc. Interface* **10**, 20120986 (2013).
30. Swahn, M. H., Braunstein, S. & Kasirye, R. Demographic and psychosocial characteristics of mobile phone ownership and usage among youth living in the slums of Kampala, Uganda. *West. J. Emerg. Med.* **15**, 600–603 (2014).
31. Basmajian, A. et al. Gendered access to digital capital and mobile phone ownership among young people in Rakai, Uganda. *Cult. Health Sex.* **25**, 1–16 (2022).
32. Campbell, B. R. et al. Mobile device usage by gender among high-risk HIV individuals in a rural, resource-limited setting. *Telemed. J. E-Health* **27**, 615–624 (2021).
33. Coston, A. et al. Leveraging administrative data for bias audits: assessing disparate coverage with mobility data for COVID-19 policy. In *Proc. 2021 ACM Conference on Fairness, Accountability, and Transparency*, 173–184 (Association for Computing Machinery, 2021).
34. Kostandova, N. et al. Adjusting mobile phone data to account for children's travel and the impact on measles dynamics in Zambia. *Am. J. Epidemiol.* <https://doi.org/10.1093/aje/kwae304> (2024).
35. Sloan, L. & Morgan, J. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. *PLoS ONE* **10**, e0142209 (2015).
36. SurveyMonkey. www.surveymonkey.com (SurveyMonkey Inc., 2024).

Acknowledgements

This work was supported by the Burroughs Wellcome Fund (Grant number 1015823.03) (AW), National Institutes of Health Director's New Innovator Award (Grant number DP2LM013102) (A.W., N.K., K.L.S., C.S., R.A., S.L., S.B.), National Institute of Allergy and Infectious Diseases (Grant number 1R01AI160780-01) (A.W., N.K., D.A.T.C., A.J.T., H.S., K.L.S., C.S., R.A., M.D.T.H., S.L., S.L., E.C.), the High Meadows Environmental Institute (HMEI) at Princeton University (B.L.R.), and National Institutes of Health (1T32AI165369-01A1) (N.K.). The funders had no role in study design, data collection, analysis, interpretation, writing of the manuscript, or decision to publish.

Author contributions

A.W. and D.A.T.C. conceptualized the study. C.S. designed the survey tool and coordinated data collection. L.R. advised on the literature search strategy. N.K., C.S., R.A., F.A., S.B., T.C., E.C., O.C.A., B.G.C., K.H.G., M.D.T.H., A.T.H., N.Ki., S.Lai, S.L.L., S.Loistate, P.M., H.R.M., R.P., T.R., J.R., B.L.R., L.R., N.R., H.S., K.L.S., A.J.T., and J.W. carried out screening and data abstraction. N.K. and A.W. carried out data cleaning and data analysis and drafted the manuscript. N.K. carried out data visualization. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54895-7>.

Correspondence and requests for materials should be addressed to Natalya Kostandova.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2024

¹Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. ²WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton SO17 1BJ, UK. ³Division of Host-Microbe Systems & Therapeutics, Department of Pediatrics, University of California San Diego, La Jolla, CA 92093, USA. ⁴Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK. ⁵Department of Biology, University of Florida, Gainesville, FL, USA. ⁶Department of Biostatistics, University of Florida, Gainesville, USA. ⁷Emerging Pathogens Institute, University of Florida, Gainesville, USA. ⁸Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, Massachusetts, USA. ⁹Program in Ecology, Evolution, and Conservation Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ¹⁰Department of Microbiology, University of Illinois Urbana-Champaign, Champaign, Illinois, USA. ¹¹Department of Statistics, University of Illinois Urbana-Champaign, Champaign, Illinois, USA. ¹²Public Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ¹³Department of Integrative Biology, University of Wisconsin-Madison, Madison, Wisconsin, USA. ¹⁴Madagascar Biodiversity Center, Antananarivo, Madagascar. ¹⁵Health Information Computation and Statistics, Lancaster Medical School, Lancaster University Lancaster, UK. ¹⁶Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. ¹⁷Informationist Services, Welch Medical Library, Johns Hopkins University, Baltimore, MD, USA. ¹⁸Population Health Sciences, Virginia Tech, Blacksburg, VA, USA. ¹⁹These authors contributed equally: Natalya Kostandova, Catherine Schluth. ²⁰These authors jointly supervised this work: Derek A. T. Cummings, Amy Wesolowski.

✉ e-mail: nkostan1@jh.edu