medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Population disruption: estimating changes in population distribution in the UK during the COVID-19 pandemic

Authors: Hamish Gibbs[1*], Naomi R Waterlow[1], James Cheshire[2], Leon Danon[3, 4, 5], Yang Liu[1], Chris Grundy[1], Adam J. Kucharski[1], LSHTM CMMID COVID-19 Working Group[†], Rosalind M. Eggo[1*]

Affiliations:

[1]Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom.

[2]Department of Geography, University College London, London, United Kingdom.

[3]Department of Engineering Mathematics, University of Bristol, Bristol, United Kingdom.

[4]The Alan Turing Institute, British Library, London, United Kingdom.

[5]Population Health Sciences, University of Bristol, Bristol, United Kingdom.

Corresponding author: [*] joint corresponding authors: hamish.gibbs@lshtm.ac.uk (HG) r.eggo@lshtm.ac.uk (RME).

[†] Membership of LSHTM CMMID COVID-19 working group is provided in the acknowledgments.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Abstract

Mobility data have demonstrated major changes in human movement patterns in response to COVID-19 and associated interventions in many countries. This can involve sub-national redistribution, short-term relocations as well as international migration. In this paper, we combine detailed location data from Facebook measuring the location of approximately 6 million daily active Facebook users in 5km^2 tiles in the UK with census-derived population estimates to measure population mobility and redistribution. We provide time-varying population estimates and assess spatial population changes with respect to population density and four key reference dates in 2020 (First lockdown, End of term, Beginning of term, Christmas). We also show how the timing and magnitude of observed population changes can impact the size of epidemics using a deterministic model of COVID-19 transmission. We estimate that between March 2020 and March 2021, the total population of the UK has declined and we identify important spatial variations in this population change, showing that low-density areas have experienced lower population decreases than urban areas. We estimate that, for the top 10% highest population tiles, the population has decreased by 6.6%. Further, we provide evidence that geographic redistributions of population within the UK coincide with dates of non-pharmaceutical interventions including lockdowns and movement restrictions, as well as seasonal patterns of migration around holiday dates. The methods used in this study reveal significant changes in population distribution at high spatial and temporal resolutions that have not previously been quantified by available demographic surveys in the UK. We found early indicators of potential longer-term changes in the population distribution of the UK although it is not clear how these changes may persist after the COVID-19 pandemic.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Introduction

Responding to the outbreak of the COVID-19 pandemic has involved the widespread use of location data collected from mobile devices[1–3]. Location data aggregated from individual GPS locations and Call Detail Records have been used as a proxy for social contact in infectious disease models[4], and have been used to measure adherence to relevant non-pharmaceutical interventions, such as domestic movement restrictions[5]. These data are typically aggregated to preserve user privacy by decreasing the spatial and/or temporal resolution of the data and applying censoring thresholds to aggregated metrics[1]. The availability of these data sources has increased as platforms including Facebook, Apple, Google, CityMapper, and mobile network providers like Vodafone and BT, have shared anonymised user data to aid the response to the pandemic[6–8].

The most common use of mobile location data for responding to COVID-19 has been as a measure of travel and activity. This data is typically aggregated, anonymised and expressed as a normalised deviation from a baseline value for specific locations or pairs of locations[1,2,5,9–11]. Personally identifiable location data are also used for contact tracing and to monitor individuals' adherence to isolation and quarantine[1,12]. In the early stages of the pandemic, mobile location data were combined with epidemiological data to estimate the spatial diffusion of SARS-CoV-2 transmission[13,14]. Research using mobile location data has demonstrated the large impact of COVID-19 and associated control measures such as movement restrictions and "stay-at-home" orders on patterns of human movement[5,10,14–16]. While there is strong evidence of disruptions in relative volumes of travel between locations, there is less clarity about the specific spatial and temporal changes in population distribution that have occurred during the pandemic.

The way that the distribution of population changes through time has important implications for public health responses, health and economic impact assessments of COVID-19, and the understanding of epidemic severity in the event of large redistributions of population. Population estimates are critical for epidemiological analysis, but research typically employs static estimates of population derived from annual or bi-annual census surveys and projections. While these static population estimates are accurate at the time of measurement, they do not reflect changes caused by major disruptions to patterns of movement and migration[17]. In the UK, there have been a limited number of surveys attempting to estimate the distribution of population during the COVID-19 pandemic, primarily focusing on specific locations like London, or specific groups such as foreign workers[18,19].

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

In this paper, we focus on creating estimates of changes in the UK population distribution at a high spatial and temporal resolution to address existing gaps in our understanding of the population distribution in the UK across continuous time scales. We propose a method for generalising the location of Facebook users to the entire UK population, and provide population estimates that "fill in" missing data between census population estimates. In the first year of the COVID-19 pandemic we observed large disruptions in the population distribution of the UK coinciding with seasonal patterns of movement and the dates of public health interventions, as well as an overall trend of decreased populations in urban areas. Coinciding with these subnational trends, we observe a decrease in the overall population of the UK. Using the estimates of populations in specific locations, we demonstrate how observed changes in population influence transmission using a model of COVID-19. In particular, we show how variations in the timing and magnitude of population changes can impact disease incidence. This has important implications for modelling approaches that employ static population estimates, as estimates may not reflect the time-varying nature of underlying populations.

Results

Population Data

Facebook, a social network and mobile application provider, records users actively sharing their location with Facebook applications, referencing these locations to an approximately 2.5km^2 gridded tile system in sequential 8-hour windows between March 10th 2020 and March 31st 2021[6]. We estimated changes in the UK population by: 1) aggregating tiles to a 5km^2 grid to align with census population estimates; 2) using 2019 mid-year population estimates for small statistical areas (Supplemental Figure 1) to estimate the spatial distribution of the UK population before the epidemic (Figure 1a); and 3) using changes in the distribution of Facebook users to adjust these population estimates each day (Figure 1b)[6,20]. The generalisation method assumes that Facebook users travel at the same rates as the rest of the UK population.

To understand the usual spatial distribution of Facebook users, we used Facebook-generated baseline estimates of population calculated in the 45 days before March 10th, from January 25th to March 9th, 2020. This baseline period is automatically back-calculated by Facebook prior to data sharing and provides the earliest available information on the distribution of Facebook users in the dataset. Before applying the

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

population adjustment, we compared the population of Facebook users in the baseline period to census population estimates across the UK to identify spatial variations in the distribution of Facebook users and census population. We found a different median percentage of Facebook users to population in each of the three 8-hour reporting periods (00:00 to 08:00, 08:00 to 16:00, and 16:00 to 0:00), of 7.46%, 9.1%, and 9.77% respectively (Figure 1c). These distributions reflect differences in the number of people actively using Facebook services at different times of day. During the baseline period, we found that the majority of tiles (77%) have percentages of Facebook usage between 5 and 20% (Supplementary Figure 2). There are some tiles where Facebook usage exceeds 75%, which was typically observed in low population tiles. Extreme discrepancies greater than 100% Facebook usage, observed in 33 tiles in total, may result from interference from sparse cellular network coverage (Supplemental Figure 3). For the remainder of the study, we used the observed number of Facebook users in tiles between March 10th, 2020 and March 31st, 2021 in the 16:00 to 00:00 window. This window has the lowest weekly variance of the number of observed Facebook users (Supplemental Figure 4), and tends to record evening/nighttime populations, making it more appropriate to compare to census population estimates which measure nighttime population.

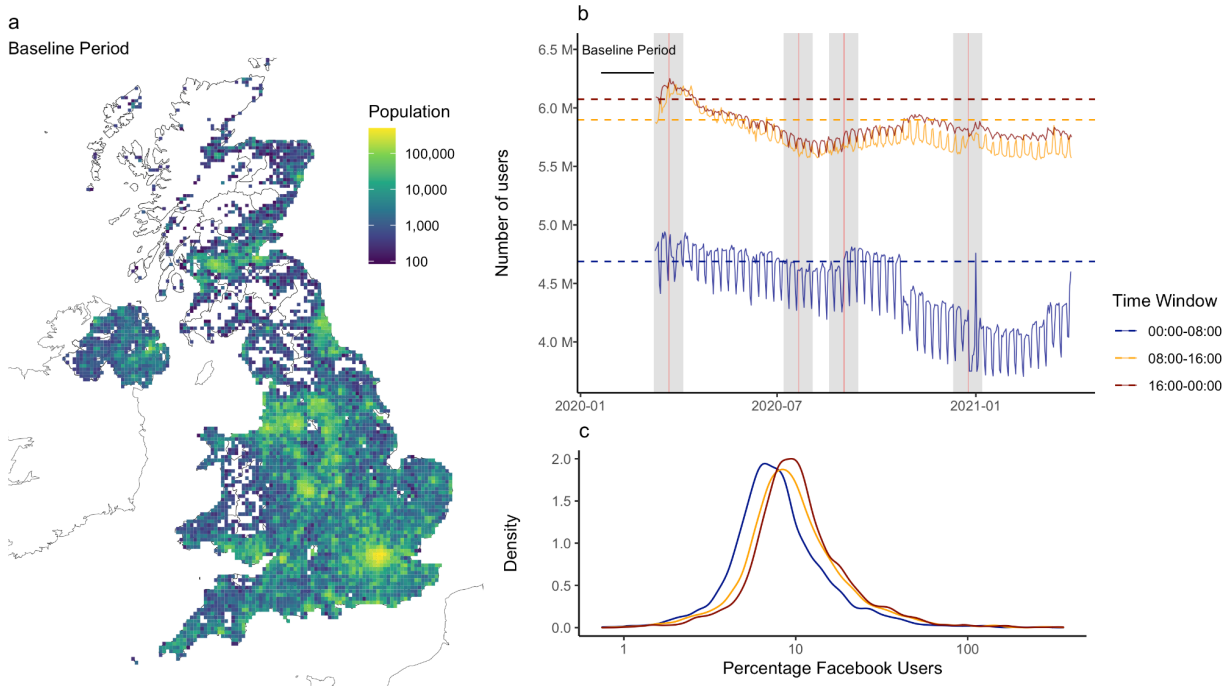

Figure 1. The relationship between Facebook users and census population estimates. *a) The population of the UK in the baseline period estimated by generalising the locations of Facebook users actively sharing location. b) The total number of Facebook users in each time window, dashed horizontal lines show baseline values used to define the proportion of users to census population. Note that axis limits for the number of users do not begin at 0. Grey bars indicate the 2 weeks preceding and following reference dates (red lines): "First national lockdown", "beginning of summer holidays", "return to school", and "Christmas". The decrease in users in the 00:00 to 08:00 period in*

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

late October most likely results from daylight savings time, which is not accounted for in these time windows. c) The distribution of the percentage of Facebook users in individual cells for each daily time window, showing that Facebook usage as a proportion of population tends to be higher in daytime periods.

Dynamic population changes

We measured spatial changes in population, focusing on four important dates in 2020: 1) the first national lockdown introduced on March 23rd, 2020; 2) the beginning of summer holidays on July 21st, 2020; 3) the return to school on September 1st, 2020; and 4) Christmas on December 25th, 2020. School start and finish dates are different in the nations of the UK and regionally, so we selected approximate values for schools in England. Note that schools in the UK were not open for in-person learning between March and July 2020 due to COVID-19 restrictions.

We observed decreases in the population of urban areas following the announcement of the first national lockdown, with a corresponding increase in less densely populated areas (Figure 2a, Supplemental Figure 5). This pattern is similar to the population changes observed in the beginning of the summer holidays period, when populations decreased in populous urban and suburban areas (Figure 2b). We observed an inverse pattern of population change during the return to school period, where populations increased in populous and decreased in less populous areas (Figure 2c). The Christmas period coincided with an announcement of travel restrictions in the South East of England on December 19th, 2020 (Tier 4), with a short term relaxation of measures on Christmas day for other areas[21] (Supplemental Figure 6). In this period, we observed a large decrease in population in central London and an increase in the population in tiles immediately peripheral to London (Figure 3d).

We calculated time-varying population estimates for the largest built-up areas (BUAs) in the UK which show major changes at key points in the epidemic (Figure 2e, Supplementary Figure 7-8). We observed a sustained decrease in the population of London between March 10th, 2020 and the implementation of public health measures on March 23rd, 2020. Comparing the 14 days before and after the first national lockdown, the population of London decreased by 3.44% (196,000 people). Further, we estimate the population of London was 646,000 lower in total between March 2020, and March 2021, with some periods of even lower population observed during the summer and Christmas holidays. Between March 2020 and March 2021, the population decreased in the 20 largest BUAs in the UK, with the largest decreases observed in Greater London, West Midlands, and Greater Manchester BUAs (Supplemental Figure 7-8).

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

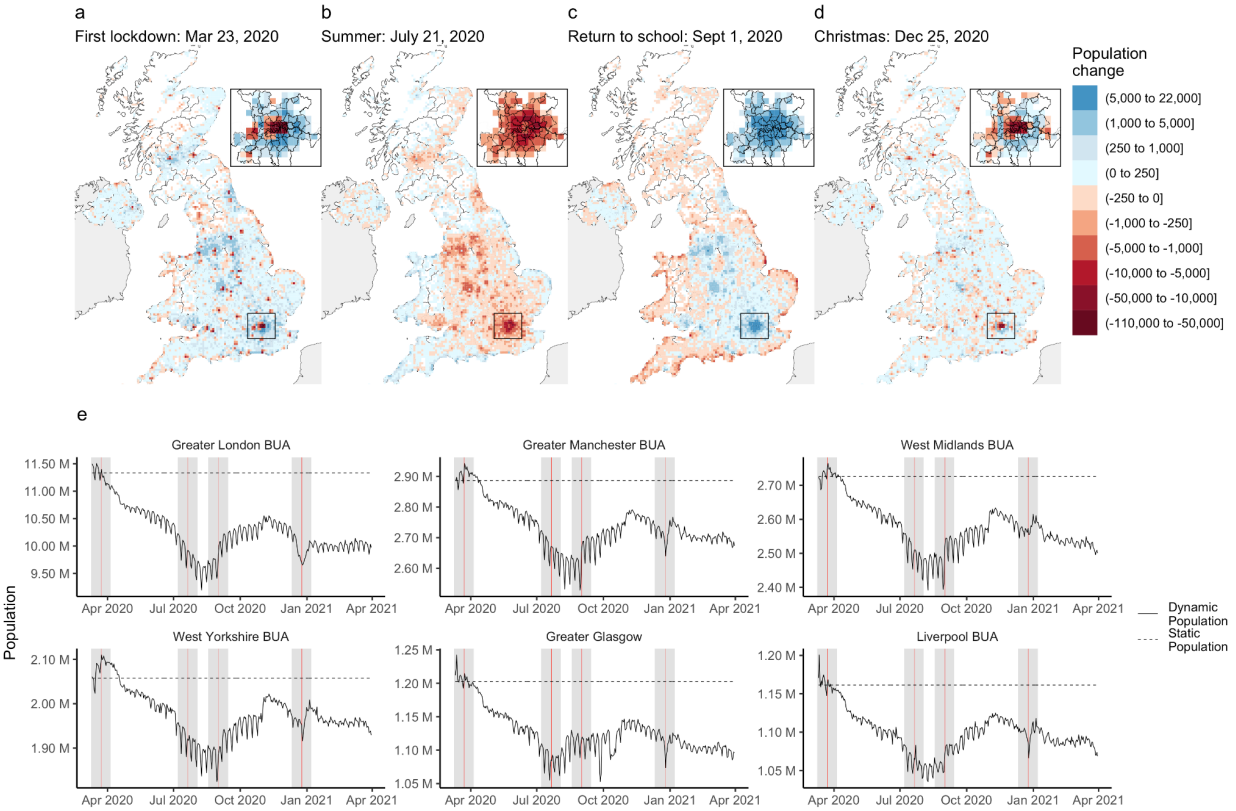

Figure 2. Time-varying estimates of population change. *Population change in the two weeks preceding and following significant reference dates: a) "First national lockdown", b) "Summer", c) "Return to school", and d) "Christmas". e) Time-varying population estimates for the six largest BUAs. Red lines indicate reference dates and grey areas show the two weeks preceding and following these dates.*

The absolute daily difference in the population size of each tile indicates the daily degree of population flux. The total absolute daily difference across the UK had strong short-term periodicity, but on a weekly level, changes in flux were clearly visible in association with the four key reference dates (Figure 3). Additionally, other national interventions also showed responses with this metric, including the second national lockdown on November 2nd 2020 marked with a 1, and the reopening of schools on March 8th (schools in the UK were shut after January 1st 2020) marked with a 2.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

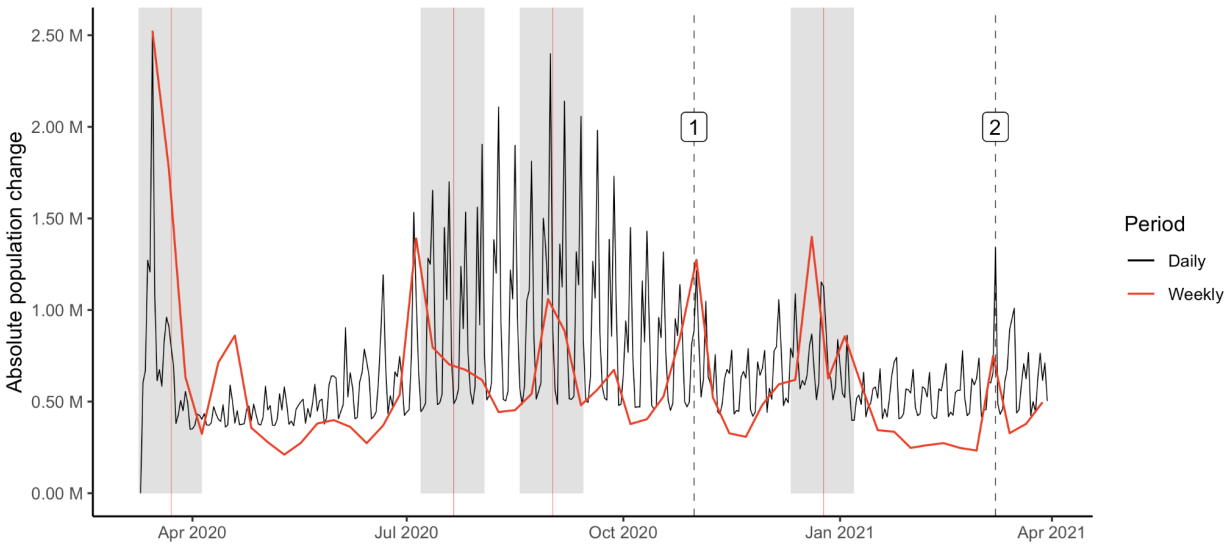

Figure 3. Total displacement of population through time. *The total change in population through time, measured as the difference between lagged population estimates in daily and weekly time windows. Vertical dashed lines show significant dates ([1] Second National Lockdown, [2] Reopening of schools).*

Rural population increase

We measured population changes in areas of different population density by dividing tiles into deciles based on their census population size (Supplemental Figure 8) and calculating time-varying population estimates for these deciles (Figure 4a). We found large decreases in the most populous decile in the first lockdown period, with the greatest increase in population observed in the 2nd most populous decile (Figure 4b). In the summer, we observed a decrease in the top two most populous deciles and a more even distribution of population among the remaining deciles, including the least populous. This reflects movement to rural areas coinciding with the beginning of summer holidays. The summer pattern is mirrored by a decrease in less populous tiles in the return to school period. Finally, we observed a similar, though reduced pattern of population decrease in the Christmas period where population in the most populous decile decreased, while in less populous deciles it increased.

Each decile showed different patterns of population change over the course of the study period (Figure 4c), where the lowest decile's population increased to 120% of census estimates during the summer period. This decile shows a smaller response to Christmas than others, with the most pronounced increases in the central deciles. Of note is that most deciles did not return to the pre-Christmas population sizes, suggesting that people could have remained where they travelled for Christmas as the UK entered new restrictions on

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

26th December. We estimated that, for the top 10% highest population tiles, the population has decreased by 6.6%.

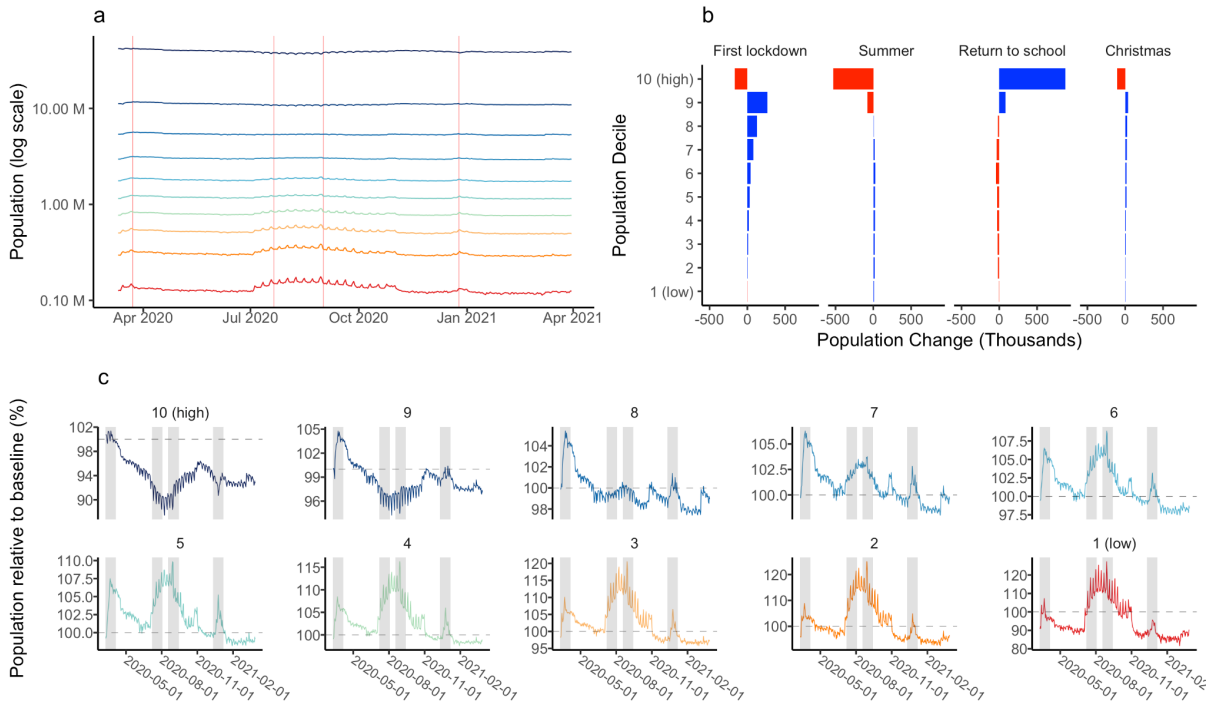

Figure 4. Population changes by population decile. *a) The national distribution of population deciles for individual tiles. b) The total population of tiles in each decile over time (log scale). c) The change in population for tiles in each population decile in the two weeks preceding and following reference dates. d) The normalized population (relative to baseline census estimate) for individual deciles through time. Note different y-axes.*

Potential impact on transmission

Deviations from census populations have implications for infectious disease dynamics and transmission models based on static population estimates. We used a simple SEIR (Susceptible-Exposed-Infectious-Recovered) model to simulate a COVID-19 epidemic while changing the underlying population size according to five different regimes of population change (Figure 5, Supplemental Figure 11).

During the study period, populations in tiles deviated from census estimates from -100% (tiles completely emptying) to greater than +1000%, with mean 0.02% and median -0.19% across all time periods (Supplemental Figure 10). Extreme outliers in population increase were rare, and reflected short increases in population in low population cells, for example, areas like beaches which experienced rare, short-term, extreme increases. Across all time periods, the distribution of deviations from census population was bimodal, with a larger number of tiles experiencing a decrease in population. The majority (95.02%) of deviations in population vary between +/- 25% for all tiles.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Changes in population size impacted the infection incidence and the final cumulative incidence (Figure 5, Supplemental Figure 11). Decreases in population of 25% at varying times relative to the progression of the epidemic decreased the epidemic curve, and the timing of the population change strongly impacted the effect on the total cumulative incidence (Fig 5a-c). For different magnitudes of population change the epidemic profile changes, and the total incidence is higher for an increased population (Fig 5d-f).

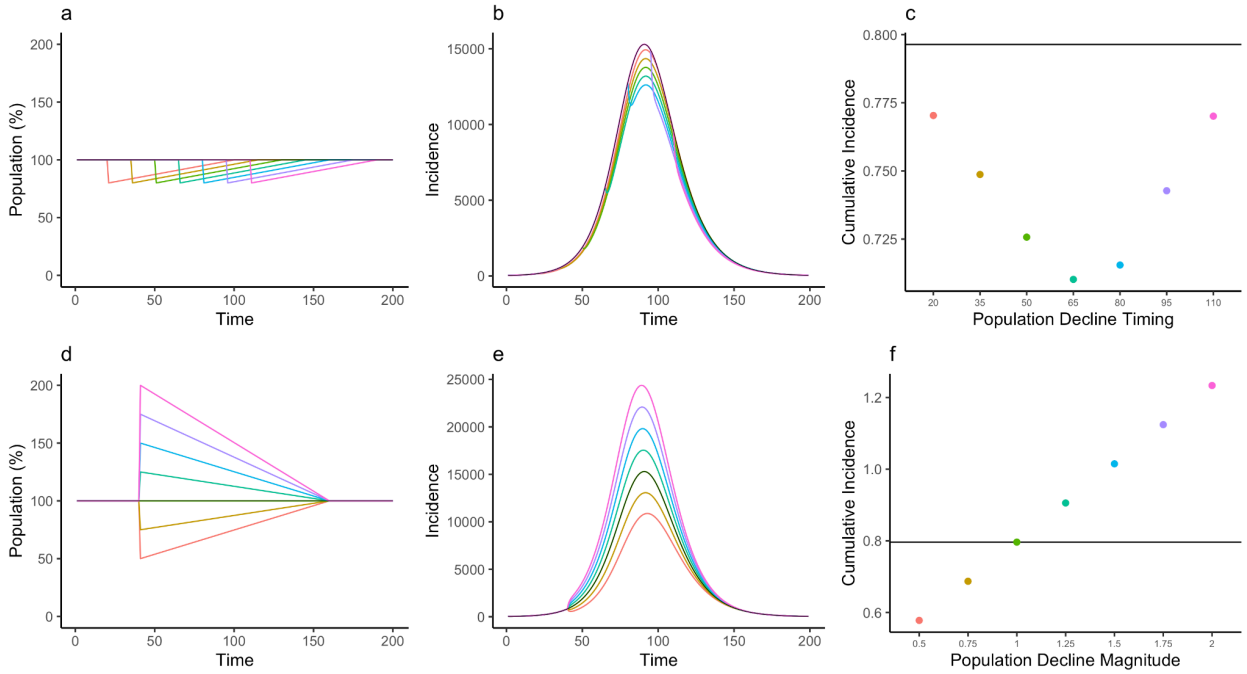

Figure 5. The effect of population change on an SEIR model of 1 million people. *a) Regimes of abrupt population decrease by 20% and slow return to baseline, starting at varying times. b) The resulting disease incidence and c) the resulting final cumulative incidence as a proportion of the population. The black line shows the "no change" scenario where population is constant. d) Regimes of population change of varying magnitude (between -50% and +100%). e) The resulting disease incidence and f) the resulting final cumulative incidence as a proportion of the original population.*

Changes in population can also affect estimates of the impact of the epidemic in real-time. Some statistics, like the rate of cases per 100,000 rely on census population estimates to calculate disease incidence relative to a population. We found that the inclusion of time-varying population estimates led to changes in the calculated COVID-19 incidence rates, leading to an altered interpretation of the severity of outbreaks in different areas. Using time-varying population estimates, we identified differences in the rate of confirmed COVID-19 cases in the 8 largest BUAs in England ranging between 23.24% and -3.26% of the rate calculated with static populations (Supplemental Figures 12-13). This difference between epidemic rates reflects changes in the underlying population of these areas through time. 84% of the time-varying rates are within 10% of the rates computed using a static

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

population. Nonetheless, the differences in epidemic rates are not trivial, particularly in urban areas during the period of highest COVID-19 cases between December 2020 and January 2021. The combined effect of high numbers of cases and population outflows from urban areas led to a consistent underestimation of the scale of the COVID-19 epidemics in the most populous Built Up Areas (Supplemental Figures 12-13).

Discussion

In this study, we present a novel approach for estimating population changes using a large, near-real time dataset of the location of millions of Facebook users in the UK. This study identifies important changes in the population of the UK coinciding with the announcement of public health interventions and with seasonal migrations in the UK. We provide evidence supporting previous reports of a decrease in population in urban areas. We also demonstrate how observed changes in population have varied in space and time. We show how these population changes can impact disease transmission and the predicted size of epidemics. This is particularly important for understanding how the timing of population changes relate to the progression of a disease epidemic. We also provide new estimates of the rate of cases using time-varying populations in different locations, which could be extended to other rates like calculated rates of vaccinations.

The COVID-19 pandemic has resulted in unprecedented changes in the population distribution of the UK. These changes include abrupt changes in population as well as broad trends of population change since the announcement of the first national lockdown in March 2020. Decreases in the population of the 20 largest urban areas in the UK have persisted throughout 2020 which may reflect new patterns of employment and home working. The trend of decreasing urban populations has not been monotonic, as the population of urban centers has increased during specific periods. Nor has the trend of population increases in less populous areas been uniform. As we demonstrate, rural populations have experienced unique population dynamics during public health interventions and seasonal migrations during summer holidays.

Our estimates of population change are similar to the limited available estimates of population changes during COVID-19 in the UK, but provide much higher spatial and temporal resolution[18,19]. In January 2021, the consulting firm PWC estimated that the population of London would decrease by 300,000 people, while a study of the non-UK born population estimated a decrease of ~600,000 people in Greater London[18,19]. While these estimates are limited, they provide helpful early evidence of the large-scale changes in

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

population that have occurred during the pandemic and reflect estimated population changes of a similar magnitude as those reported in detail in this study (646,000 in Greater London).

The population estimates calculated in this study are a novel resource for understanding the evolution of the UK population through time and could be used to “fill the gap” between census population estimates. In the future, these estimates can be compared to results from the 2021 census to gain a detailed understanding of their accuracy at a high spatial resolution, and to identify potential biases in the generalization method. It is important to note that demographic trends are informed by varying rates of births, deaths, immigration and emigration. Changes in commuting patterns, the labor market, and seasonal migrations also impact patterns of population change[17,19]. The estimates presented in this study are intended to provide detailed information on dynamic population changes between March 2020 and March 2021 but it is unclear how the trends of population change observed during the COVID-19 pandemic will persist in the future.

We demonstrate how transmission dynamics can be influenced by dynamic changes in population size using a transmission model of COVID-19. While extreme increases in population are rare, the regular fluctuation of populations through time is an important factor which is not widely incorporated in epidemiological models, and something that occurs at a frequency shorter than standard intervals for census population estimation. We also demonstrated that time-varying population estimates impact reported rates of disease incidence. Other measurements of disease impact (attack rate, hospitalisation rate) could also be calculated by incorporating contemporaneous (time-varying) estimates of population. While this information is useful for retrospective reinterpretations of the severity of disease in particular locations, it would be particularly valuable for the modelling of disease spread, where time-varying population estimates could more accurately reflect population distributions in real time.

This study and the proposed method of population generalisation have a number of limitations. As the basis for our population generalisation, we use the closest available census population estimates to the baseline period. As demonstrated in this study, the population of the UK experiences dynamic redistributions and it is not possible to identify any population changes which occured between the time of census estimation and the baseline period. The estimates presented in this study will be valuable for comparison to results from the 2021 census, and can provide further information on the use of alternative sources of population data for measuring patterns of population change.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Further research is required to fully understand the demographic characteristics of Facebook users who are presented in aggregated population and mobility metrics, and how the behaviour of these individuals varies from the general population[22,23]. There is still a limited understanding of how user subsets from applications like Facebook vary from the general population and how this difference may be reflected in aggregated location metrics. In the future, research on the bias of these user subsets could be used to improve the generalisation of the behaviour of these individuals for representing the entire population.

Conclusion

Time-varying population estimates provide detailed information on major changes in the population distribution of the UK during the COVID-19 pandemic. Generalising the movement of Facebook users, we present strong evidence of population decreases in major urban areas in the UK and demonstrate the ways that these changes impact transmission models.

Methods

Data

We used population data provided by the Facebook Data for Good program[6] which records the number of active Facebook users in spatial tiles in sequential time periods. We used data from March 10th 2020 to March 31st 2021, and we also rely on the Facebook-generated baseline from a 45 day period between January 29th and March 9th 2020 as our reference for the baseline population of Facebook users.

Facebook datasets are provided in a gridded spatial reference, referred to as “tiles” which are referenced by a unique integer (a “quadkey”). The resolution of tiles is defined by a “zoom level.” The population data is provided at zoom level 13 (approximately 2.5 km^2) and zoom level 12 (approximately 5 km^2). We aggregate population data to zoom level 12 to combine it with census population estimates. There are 4 level 13 tiles for each level 12 tile. Data are provided in sequential 8-hour periods (00:00 - 08:00, 08:00 - 16:00, and 16:00 to 00:00) and are available with approximately 48 hours delay. Baseline population values are defined as the median number of users in a tile for each 8-hour period during a baseline period. These values provide an aggregated measure of the population of individual tiles,

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

which we use to incorporate information preceding the start of the full dataset on March 10th 2020.

We applied a population adjustment to the population of Facebook users, transforming the number of Facebook users into the number of actual people using 2019 mid-year population estimates from small statistical areas in the UK (England and Wales: Output Areas, Northern Ireland: Small Areas, Scotland: Data Zones). These estimates are the highest resolution population data produced by UK statistics agencies. We used 2019 mid-year estimates as they are the closest estimates available to the baseline period of the Facebook data. There is uncertainty in the relationship between 2019 population estimates and the distribution of Facebook users due to unobserved changes in population between 2019 and March 2020.

Facebook population data do not include demographic information and therefore, we are not able to compare to the total UK population on age structure, gender, ethnicity, or socioeconomic status because of the privacy-preserving structure of the dataset. In previous research, we have found no associations with key demographic factors at the tile level[24]. Users may also be removed from the dataset due to varying patterns of usage including deactivating their accounts or pausing their use of Facebook services, or because of international travel. Finally, because location sharing is an "opt-out" feature for Facebook users, there may be bias in the population of Facebook users included in these datasets. We use data from a specific time period (16:00 to 00:00) to reduce the influence of varying patterns of usage.

Population Adjustment

We estimated the population from the number of Facebook users by combining Facebook population data with census population estimates. First, we extracted census Population estimates to zoom level 12 tiles by assigning the population weighted centroid of each census to the overlapping tile. We then summed the census population estimates for each tile

$$P_i = \sum_{j=1}^{n} O_{i,j}$$

where P_i is the census population of the i^{th}tile, and $S_{i,j}$is the j^{th}OA population estimate intersecting the ith tile. We then computed the the median number of Facebook users in each tile in the baseline period for each tile

$$F_i = median(F_{i,j}) \text{ for } i = 1,\dots, n$$

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

where F_i is the reference number of Facebook users in the ith tile and $F_{i,b}$ is the jth baseline value for the the ith tile. We then computed the proportion of Facebook users to census population for each tile

$$Prop_i = \frac{F_i}{P_i}$$

where the $Prop_i$ is the proportion of Facebook users, F_i is the baseline number of Facebook users, and P_i is the census population estimate of the ith tile. We then adjusted the number of Facebook users to the population using these proportions

$$P_{adj_i} = \frac{F_{obs_i}}{Prop_i}$$

where P_{adj_i} is the adjusted population, F_{obs_i} is the observed number of Facebook users, and $Prop_i$ is the baseline proportion of Facebook users to census population for the ith tile.

Population Variation

We measured the change in population in individual cells in sequential daily and weekly time periods as a proxy for population movement. To compute this measurement in daily periods, we computed the difference between lagged and current population estimates for individual tiles. We then summed the absolute value of these differences for daily periods.

To compute the difference in population change for weekly periods, we computed the average population of individual tiles per week, then calculated the difference between lagged and current population estimates for individual tiles. We summed the absolute value of these differences for weekly periods to measure weekly changes in population.

Population Deciles

We assigned tiles to population deciles based on the value of their census populations. This labelling was consistent across the time series and was not altered in response to population changes in individual tiles.

Transmission Model

To simulate the progression of a sample epidemic in the presence of changes in population size, we constructed a SEIR model which models the number of individuals in Susceptible (S), Exposed (E), Infectious (I), Recovered (R) compartments through time. We assumed population changes occurred independently of the infectious state. The progression of the

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

modelled disease is influenced by the parameter R_0, which defines the average number of infections that will result from a single infection in a completely susceptible population, δ, the latent period, and γ, the duration of infectiousness. In our example, we use a model where $R_0 = 2$, $\delta = \frac{1}{3}$, $\gamma = \frac{1}{5}$[25,26], initial population = 1,000,000.

Incidence Rates

We calculated the rate of COVID-19 cases using data on confirmed COVID-19 cases by specimen date in Lower Tier Local Authorities (LTLAs)[27]. We aggregated this data to Built Up Areas (BUAs) using a lookup of LTLAs to BUAs[28]. We restricted this method to the top 20 BUAs by population (total BUAs: 6,347) because these BUAs encompassed multiple LTLAs, ensuring spatial consistency between COVID-19 case data and population data.

We then extracted census and dynamic (estimated) population data from tiles to BUAs. We computed the rate of COVID-19 cases per 100k people against a static population for each time period using

$$Rate_{static} = \left(\frac{Cases}{Pop_{static}}\right) * 100000$$

Where $Rate_{static}$ is the rate of COVID-19 cases $Cases$ in a given time period against a static population Pop_{static}. We computed the same rate using a time-varying (dynamic) population using:

$$Rate_{dynamic} = \left(\frac{Cases}{Pop_{dynamic}}\right) * 100000$$

Where $Rate_{dynamic}$ is the rate of COVID-19 cases $Cases$ in a given time window by population, $Pop_{dynamic}$.

Data and Code Availability

The terms of use of data from the Facebook Data for Good program prohibit unauthorised distribution. Data is available from the Facebook Data for Good Partner Program by application. Data processing and analysis code is available from https://github.com/cmmid/uk_fb_population_estimation. We have also aggregated our population estimates to 2019 Local Authority Districts, available from https://zenodo.org/record/5013620.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Acknowledgements

The following authors were part of the Centre for Mathematical Modelling of Infectious Disease COVID-19 Working Group. Each contributed in processing, cleaning and interpretation of data, interpreted findings, contributed to the manuscript, and approved the work for publication: Damien C Tully, Alicia Rosello, David Hodgson, Kaja Abbas, Gwenan M Knight, Sam Abbott, Katherine E. Atkins, Christopher I Jarvis, Kiesha Prem, Matthew Quaife, Graham Medley, Stefan Flasche, Frank G Sandmann, Rosanna C Barnard, William Waites, Nikos I Bosse, Rachel Lowe, C Julian Villabona-Arenas, Simon R Procter, W John Edmunds, Mark Jit, Ciara V McCarthy, Mihaly Koltai, Akira Endo, Billy J Quilty, Rachael Pung, Oliver Brady, Amy Gimma, Timothy W Russell, Carl A B Pearson, Stéphane Hué, Fiona Yueqian Sun, Joel Hellewell, James D Munday, Emilie Finch, Samuel Clifford, Katharine Sherratt, Paul Mee, Nicholas G. Davies, Lloyd A C Chapman, Kathleen O'Reilly, Sophie R Meakin, Yalda Jafari, Kerry LM Wong, Sebastian Funk.

Funding

The following funding sources are acknowledged as providing funding for the named authors. This research was partly funded by the Bill & Melinda Gates Foundation (INV-003174: YL). EDCTP2 (RIA2020EF-2983-CSIGN: HPG). This project has received funding from the European Union's Horizon 2020 research and innovation programme - project EpiPose (101003688: YL). HDR UK (MR/S003975/1: RME). This research was partly funded by the National Institute for Health Research (NIHR) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care (16/137/109: YL; NIHR200908: AJK, RME). UK DHSC/UK Aid/NIHR (PR-OD-1017-20001: HPG). UK MRC (MC_PC_19065 - Covid 19: Understanding the dynamics and drivers of the COVID-19 epidemic using real-time outbreak analytics: RME, YL). Wellcome Trust (206250/Z/17/Z: AJK). UK MRC (MC_PC_19067, MR/V038613/1 LD), UK EPSRC (EP/V051555/1: LD), The Alan Turing Institute under the EPSRC (EP/N510129/1: LD).

The following funding sources are acknowledged as providing funding for the working group authors. This research was partly funded by the Bill & Melinda Gates Foundation

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

(INV-001754: MQ; INV-003174: KP, MJ; INV-016832: SRP; NTD Modelling Consortium OPP1184344: CABP, GFM; OPP1139859: BJQ; OPP1191821: KO'R). BMGF (INV-016832; OPP1157270: KA). CADDE MR/S0195/1 & FAPESP 18/14389-0 (PM). ERC Starting Grant (#757699: MQ). ERC (SG 757688: CJVA, KEA). This project has received funding from the European Union's Horizon 2020 research and innovation programme - project EpiPose (101003688: AG, KLM, KP, MJ, RCB, WJE). FCDO/Wellcome Trust (Epidemic Preparedness Coronavirus research programme 221303/Z/20/Z: CABP). This research was partly funded by the Global Challenges Research Fund (GCRF) project 'RECAP' managed through RCUK and ESRC (ES/P010873/1: CIJ). HPRU (This research was partly funded by the National Institute for Health Research (NIHR) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care200908: NIB). MRC (MR/N013638/1: EF; MR/V027956/1: WW). Nakajima Foundation (AE). NIHR (16/136/46: BJQ; 16/137/109: BJQ, FYS, MJ; 1R01AI141534-01A1: DH; NIHR200908: LACC; NIHR200929: CVM, FGS, MJ, NGD; PR-OD-1017-20002: AR, WJE). Royal Society (Dorothy Hodgkin Fellowship: RL). Singapore Ministry of Health (RP). UK MRC (MC_PC_19065 - Covid 19: Understanding the dynamics and drivers of the COVID-19 epidemic using real-time outbreak analytics: NGD, SC, WJE; MR/P014658/1: GMK). UKRI (MR/V028456/1: YJ). Wellcome Trust (206250/Z/17/Z: TWR; 206471/Z/17/Z: OJB; 208812/Z/17/Z: SC, SFlasche; 210758/Z/18/Z: JDM, JH, KS, SA, SFunk, SRM; 221303/Z/20/Z: MK). No funding (DCT, SH).

References

1. Grantz, K. H. *et al.* The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat. Commun.* **11**, 4961 (2020).
2. Oliver, N. *et al.* Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci. Adv.* **6**, eabc0764 (2020).
3. Mobile Location Data and Covid-19: Q&A. *Human Rights Watch* https://www.hrw.org/news/2020/05/13/mobile-location-data-and-covid-19-qa (2020).
4. Tizzoni, M. *et al.* On the Use of Human Mobility Proxies for Modeling Epidemics. *PLOS Comput. Biol.* **10**, e1003716 (2014).
5. Jacobsen, G. D. & Jacobsen, K. H. Statewide COVID-19 Stay-at-Home Orders and Population Mobility in the United States. *World Med. Health Policy* **12**, 347–356 (2020).
6. Disease Prevention Maps. *Facebook Data for Good* https://dataforgood.fb.com/tools/disease-prevention-maps/.
7. COVID-19 Community Mobility Report. *COVID-19 Community Mobility Report* https://www.google.com/covid19/mobility?hl=en.
8. COVID-19 - Mobility Trends Reports. *Apple* https://www.apple.com/covid19/mobility.
9. Overview - Community Mobility Reports Help. https://support.google.com/covid19-mobility/answer/9824897?hl=en-GB.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

10. Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S. & Colizza, V. Population mobility reductions during COVID-19 epidemic in France under lockdown. *medRxiv* 2020.05.29.20097097 (2020) doi:10.1101/2020.05.29.20097097.
11. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study - The Lancet Infectious Diseases. https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30553-3/fulltext.
12. Applications of digital technology in COVID-19 pandemic planning and response - The Lancet Digital Health. https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30142-4/fulltext.
13. Quilty, B. J. *et al.* The effect of travel restrictions on the geographical spread of COVID-19 between large cities in China: a modelling study. *BMC Med.* **18**, 259 (2020).
14. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
15. Gibbs, H. *et al. Changing travel patterns in China during the early stages of the COVID-19 pandemic*. http://medrxiv.org/lookup/doi/10.1101/2020.05.14.20101824 (2020) doi:10.1101/2020.05.14.20101824.
16. Pepe, E. *et al.* COVID-19 outbreak response: a first assessment of mobility changes in Italy following national lockdown. *medRxiv* 2020.03.22.20039933 (2020) doi:10.1101/2020.03.22.20039933.
17. National Population Projections Accuracy Report - Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/methodologies/nationalpopulationprojectionsaccuracyreport.
18. London population set to decline for first time since 1988 – report. *the Guardian* http://www.theguardian.com/uk-news/2021/jan/07/london-population-decline-first-time-since-1988-report-covid-home-working (2021).
19. Estimating the UK population during the pandemic. *ESCoE* https://www.escoe.ac.uk/estimating-the-uk-population-during-the-pandemic/ (2021).
20. Population estimates for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019estimates.
21. Prime Minister announces Tier 4: 'Stay At Home' Alert Level in response to new COVID variant. *GOV.UK* https://www.gov.uk/government/news/prime-minister-announces-tier-4-stay-at-home-alert-level-in-response-to-new-covid-variant.
22. Ribeiro, F. N., Benevenuto, F. & Zagheni, E. How Biased is the Population of Facebook Users? Comparing the Demographics of Facebook Users with Census Data to Generate Correction Factors. in *12th ACM Conference on Web Science* 325–334 (Association for Computing Machinery, 2020). doi:10.1145/3394231.3397923.
23. Lansley, G. & Cheshire, J. Challenges to representing the population from new forms of consumer data. *Geogr. Compass* **12**, e12374 (2018).
24. Gibbs, H. *et al.* Detecting behavioural changes in human movement to inform the spatial scale of interventions against COVID-19. *medRxiv* 2020.10.26.20219550 (2021) doi:10.1101/2020.10.26.20219550.
25. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period | Science. https://science.sciencemag.org/content/368/6493/860.
26. Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, (2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.06.22.21259336; this version posted June 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

27. Download data | Coronavirus in the UK. https://coronavirus.data.gov.uk/details/download.

28. Statistics, O. for N. Built-up Areas (December 2011) Boundaries V2. https://data.gov.uk/dataset/15e3be7f-66ed-416c-b0f2-241e87668642/built-up-areas-december-2011-boundaries-v2 (2017).