Quantifying the dynamics of migration after Hurricane Maria in Puerto Rico

Rolando J. Acosta[a,1] ⓘ, Nishant Kishore[b,1] ⓘ, Rafael A. Irizarry[a,c,2], and Caroline O. Buckee[b,2,3] ⓘ

[a]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA 02115; [b]Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115; and [c]Department of Data Science, Dana-Farber Cancer Institute, Boston, MA 02215

Population displacement may occur after natural disasters, permanently altering the demographic composition of the affected regions. Measuring this displacement is vital for both optimal postdisaster resource allocation and calculation of measures of public health interest such as mortality estimates. Here, we analyzed data generated by mobile phones and social media to estimate the weekly island-wide population at risk and within-island geographic heterogeneity of migration in Puerto Rico after Hurricane Maria. We compared these two data sources with population estimates derived from air travel records and census data. We observed a loss of population across all data sources throughout the study period; however, the magnitude and dynamics differ by the data source. Census data predict a population loss of just over 129,000 from July 2017 to July 2018, a 4% decrease; air travel data predict a population loss of 168,295 for the same period, a 5% decrease; mobile phone-based estimates predict a loss of 235,375 from July 2017 to May 2018, an 8% decrease; and social media-based estimates predict a loss of 476,779 from August 2017 to August 2018, a 17% decrease. On average, municipalities with a smaller population size lost a bigger proportion of their population. Moreover, we infer that these municipalities experienced greater infrastructure damage as measured by the proportion of unknown locations stemming from these regions. Finally, our analysis measures a general shift of population from rural to urban centers within the island. Passively collected data provide a promising supplement to current at-risk population estimation procedures; however, each data source has its own biases and limitations.

Hurricane Maria | Puerto Rico | population displacement | passively collected data

In the aftermath of natural disasters, both short-term population displacement and longer-term migration may occur, leaving some affected regions permanently altered (1–3). Measuring population displacement is a priority in the immediate days and weeks after an event like a hurricane or flood for the provision of aid and other supplies to communities in need. It is also critical to inform mortality estimates and other measures of public health interest that require an up-to-date denominator since census estimates may be rendered inaccurate (4, 5). Measuring shifts in demographic composition and the geographic distribution of populations on longer timescales is also critical to rebuilding efforts and to the development of frameworks for building resilience to future disasters. Currently, however, few data sources can be used to rapidly assess and monitor population displacement in the short- and medium-term timescales after disasters happen (6).

In the absence of reliable migration data in the wake of natural disasters, the population size estimate used by government agencies and researchers generally relies on census estimates and assumes a linear change in population size between intervals or a constant population size since the most recent estimate (7, 8). New approaches to estimating fluctuating denominators in near real time would greatly improve disaster response and the assessment of local needs in the short- and long-term aftermath. Rapid censuses conducted in short intervals before and after a disaster are both logistically and financially impractical. In an increasingly digitally connected world, however, passively collected digital records are often maintained by technological services providers for billing or marketing purposes. These data, such as flight information, mobile phone data, or social media traces, can provide insight into the fluctuation of their respective populations at a high temporal and geographic resolution, both before and after a disaster. Assuming that appropriate steps are taken to anonymize and aggregate these data streams in secure ways, novel data streams offer ways to assess the needs of populations more accurately (9, 10).

Hurricane Maria made landfall in Puerto Rico as a category 4 storm on 20 September 2017, becoming the third costliest hurricane in US history (11, 12). In the ensuing weeks, the damage to infrastructure caused by the storm resulted in a widespread lack of access to electricity, communication, and health services (13, 14). Population displacement off the island and within Puerto Rico was widespread, although this was difficult to monitor directly. Direct and indirect mortality caused by the storm also increased in the months after the hurricane (8, 13–17), but estimating mortality was made more complicated by the migration of populations because the population at risk in different parts of the island was shifting over time. Due to Hurricane Maria, the US Census Bureau ceased operations of the Puerto Rico Community Survey (PRCS; a monthly survey of 36,000 housing units across the island) from October to December 2017. Operations resumed in January 2018, with early results showing an island-wide increase in

Significance

Understanding the population composition and distribution of a region affected by a major natural disaster is vital for the allocation of resources to communities in need and critical to inform mortality estimates. Currently, the US Census Bureau is the only institution that publishes reliable population estimates for the United States and its territories. Since these are published once per year, it is impossible to use census-based population estimates to assess short-term postdisaster out-of-jurisdiction migration and within-jurisdiction migration. The utilization of social media traces, coupled with mobile phone data, could provide live estimates of postdisaster population changes in disaster-affected areas.

out migration in 2017 compared with 2016 (18). The most recent updates from the PRCS show a continued increase in out migration with a general population that decreased by 4% from 2017 to 2018 (19).

In this study, we evaluate two passively collected data sources and compare them with data on air travel and census data to evaluate the effects of Hurricane Maria on large-scale population fluctuation in Puerto Rico. We investigate the ability of these data sources to estimate the island-wide population at risk postdisaster and identify within-island geographic heterogeneity in population migration after the storm. We observe a nonlinear change in population at risk following Maria and show that this difference is affected by rurality. These passively collected data sources also provide insight into regions that are more heavily affected by disasters and can augment the resources available to first responders. Each data source has its own limitations and biases and could be used in conjunction with traditional census-based population estimates to improve the response to natural disasters, and to understand how to build more resilient systems in anticipation of future events.

Results

Data Sources. We compared four independent datasets to estimate population changes over the course of a year after the hurricane. First, we obtained the intercensal yearly population estimates provided by the American Community Survey (ACS) for 2010 to 2018, a yearly survey conducted by the US Census Bureau. The ACS estimates are for 1 July of each year. We considered this estimate to be our gold standard.

Second, we extracted data from Disaster Maps provided by Facebook's Data for Good team. Specifically, for a group of people determined to be living in Puerto Rico the week before Hurricane Maria, weekly estimates from 21 August 2017 to 30 July 2018 on the number of these users residing in 77 of the 78 municipalities on the island were available. Note that these data are a closed cohort of known individuals, so no new users appear in the data, and there is a constant rate of attrition of users expected. However, this is also the only dataset on a municipality, rather than an island-wide, level.

Third, Teralytics provided island-wide daily population proportion estimates from 31 May 2017 to 30 April 2018 based on cell phone usage patterns, analyzed in partnership with an unnamed mobile operator on the island. Since we do not know which operator the data are from, it was impossible to assess the geographic bias in ownership in this group. The proportions were calculated relative to a baseline population determined by Teralytics using a method unavailable to the authors. Due to the unreliability of data for the 4 wk after Hurricane Maria, presumably due primarily to low connectivity, proportion estimates were not provided for this time period.

Finally, we obtained Airline Passenger Traffic (APT) data from the US Bureau of Transportation Statistics (BTS) through the Puerto Rico Institute of Statistics. The data are composed of monthly counts of passengers who arrived and left the island per month from January 2010 to February 2018. These data are unbiased, but they have a coarse temporal and geographic scale and cannot account for the same individuals on repeated trips or distinguish Puerto Rican residents from short-term aid workers and other visitors.

Details on how each of these datasets was constructed are included in *Methods*, and a side-by-side comparison is found in *SI Appendix*, Table S1.

Population Decreased after Hurricane Maria. We found agreement across all data sources of a consistent loss of population from 1 July 2017 to 1 July 2018; however, the dynamics and magnitude of the loss differ between data types (Fig. 1). The ACS predicts a population loss of 129,848, a 4% decrease; the APT data predict a population loss of 168,295, a 5% decrease; the Teralytics data predict a population loss of 235,375, an 8% decrease; and the Facebook data predict a 17% decrease, which equates to a total estimated population loss of 475,779 on the island. In all cases, we observed a sharp drop after the hurricane. Both the APT data and Teralytics data show a rebound in population and stabilization after 31 December 2017. The Facebook data do not show stabilization until April 2018. We note that Facebook data are based on a closed cohort that is not able to measure population increases due to immigration to Puerto Rico. However, it does represent the cohort that likely experienced Hurricane Maria since the selection criteria used by Facebook's algorithm ensure that transient populations, such as tourists, are not included in the closed cohort.
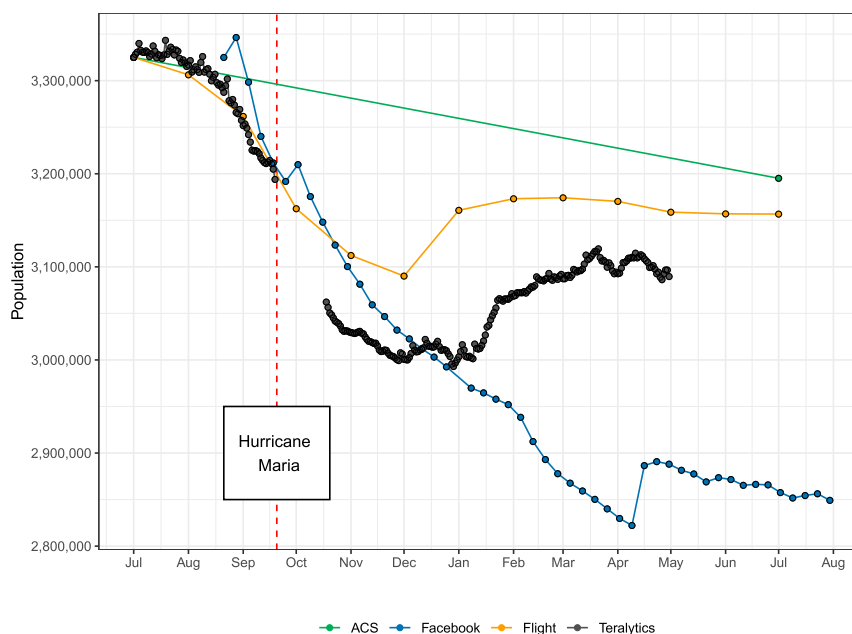
**Fig. 1.** Population estimates for each data source.

Small Municipalities Lost a Larger Portion of Their Population Compared with Large Ones. Facebook provided municipality-level data that allowed us to assess within-island migration. For each municipality, the data included weekly counts of individuals still in the municipality and new to the municipality. For privacy protection, the new to municipality data were not included if the number of users was below a prespecified threshold (*Methods* has details on the data imputation techniques used, as well as a sensitivity analysis comparing different imputation approaches). We found that the share of users belonging to each municipality was highly correlated with the baseline population size of each municipality (*SI Appendix*, Fig. S1). On average, municipalities with a smaller population size lost a bigger proportion of their population during the study period (Fig. 2). The municipality of Toa Baja was a large outlier (*SI Appendix*, Fig. S2) due to a reported surge of new individuals moving there at the end of the study period. Possible explanations for Toa Baja being an outlier are provided in *Discussion*. Apart from this outlier, San Juan, which is the capital of Puerto Rico, is the only municipality with more individuals at the end of the study period relative to baseline.

Within-Island Migration Shifted Populations from Rural to Urban Regions. All municipalities experienced loss of their baseline resident populations. Much of this loss is explained by off-island immigration (Fig. 3). However, the locality gaining the most new residents was San Juan (Fig. 3). In fact, for urban areas, the baseline resident population loss was compensated by in migration from other municipalities. For example, San Juan experienced a 19% loss of its baseline resident population but ended the study period with a 7% increase in total population, suggesting an in migration of 26% by the end of the study period. Another appealing aspect of these data is that they provide information on the destination of those displaced (Fig. 3*B*). Note that the top two destinations for Facebook users in our cohort were Miami, FL, and Jacksonville, FL, respectively. These results taken together suggest a migration pipeline from rural to urban municipalities and likely off island.

Infrastructure Damage Was Greater in Rural Areas. Disaster Maps provided information about the proportion of people whose location is unknown for a particular week. Unknown locations may be the result of 1) people stopping use of Facebook, 2) people
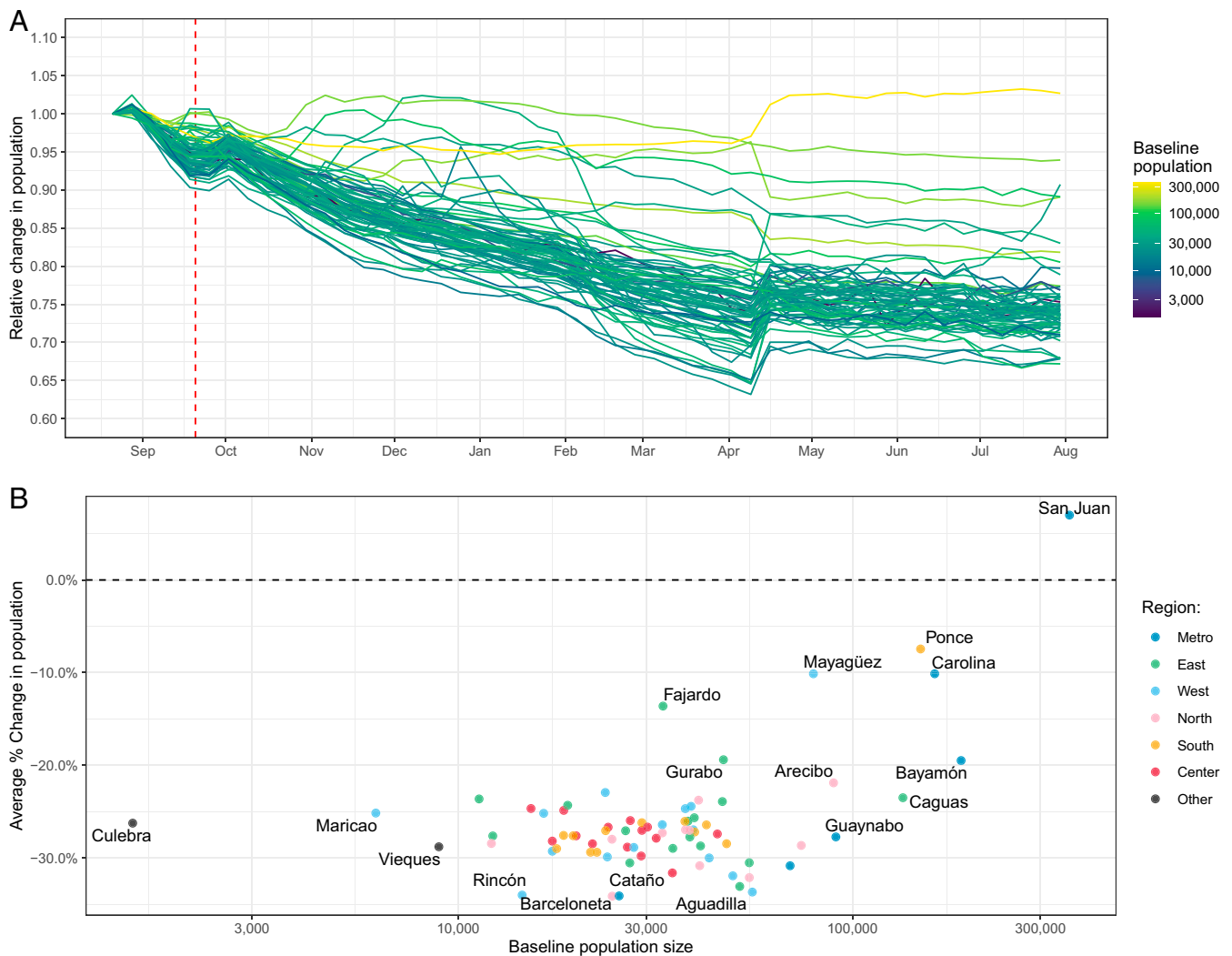
Fig. 2. Small municipalities lost a larger portion of their population compared with large ones. (*A*) Relative change in the population of Facebook cohort members per municipality. Each curve corresponds to a different municipality. Population size in July 2017 is denoted by color. (*B*) Average percentage change in population at the end of the study period, relative to baseline, compared with the ACS population size of the municipality at baseline. The geographical regions are presented in *SI Appendix*, Fig. S5. We fitted a linear model to each population curve and computed the average percentage change using the first and last fitted values. The vertical line corresponds to the day when Hurricane Maria made landfall in Puerto Rico.
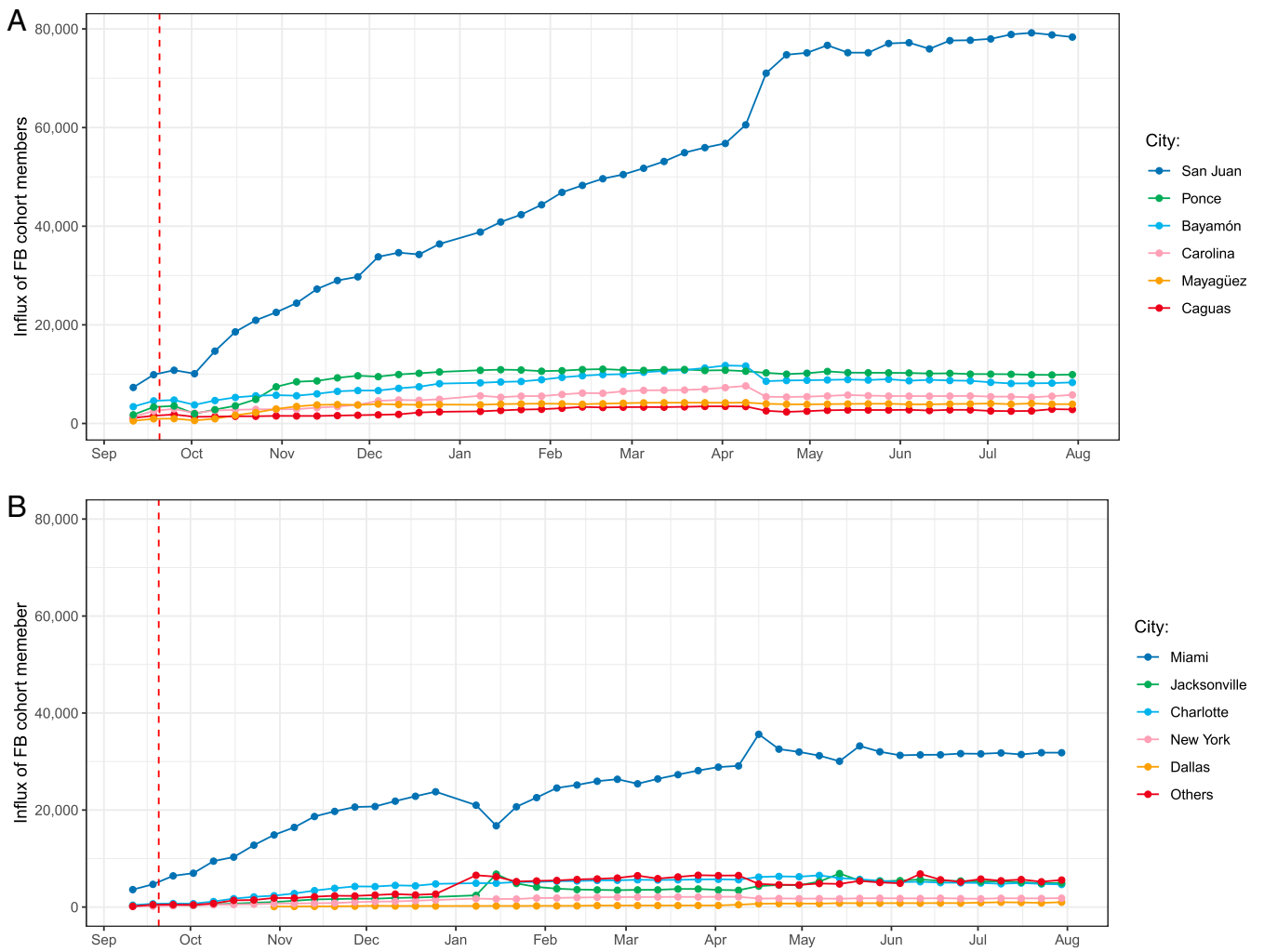
Fig. 3. Immigration of Puerto Rican Facebook (FB) cohort members within and outside of Puerto Rico. (*A*) Top six municipalities in Puerto Rico in terms of influx of FB cohort members who were located elsewhere before Hurricane Maria. (*B*) Top five destinations of FB cohort members after Hurricane Maria. The "Others" curve corresponds to the influx of all other destinations together. The vertical lines correspond to the day when Hurricane Maria made landfall in Puerto Rico.

having changed their Facebook behavior, or 3) loss of electricity and communication infrastructure. In the first full week of data collection immediately after the hurricane, approximately half of the cohort members did not register a location (Fig. 4*A*), likely due to the widespread loss of infrastructure. During this week, the proportion of users with reported unknown locations was substantially higher in the rural areas compared with in urban areas (Fig. 4*B* and *SI Appendix*, Fig. S3). We focused our analysis to only include the week immediately following the impact of Hurricane Maria to ensure that any loss of individuals in a region was not likely to be due to population migration. This is especially important as there was widespread damage to roads and loss of transportation infrastructure during this time (12).

In *SI Appendix*, Table S2, we show the top 15 municipalities that had a greater proportion of cohort members with their location unknown compared with the island-wide average. While rurality is directly linked with availability of infrastructure, the Facebook data highlight the heterogeneity in loss of access to resources in the immediate aftermath of a disaster. Some municipalities continued to have missing data relative to the island-wide average for many months.

Discussion

We show that passively collected data sources for estimating population displacement may provide insights into the dynamics

of migration following a natural disaster that cannot be obtained using traditional methods. Our results point to a consistent and long-term loss of population in Puerto Rico after Hurricane Maria. In the Teralytics data and the flight data, the decrease in population levels off in December 2017, and in January 2018, we see a rebound in Puerto Rico's population. In the Facebook data, the population loss continues until April 2018, resulting in the lowest overall population estimate. Both datasets provided island-wide population estimates that were lower than the ACS vintage 2018 point estimate. The discrepancies in population estimates are explained, in part, by the different sources. For example, the Facebook data constitute a closed cohort, whereas the Teralytics and flight data are two open cohorts with different inclusion and exclusion dynamics, and ACS data are based on surveys carried out on the island (*SI Appendix*, Table S1). Nevertheless, taken together, these data sources highlight the dynamic and consistent population displacement in the wake of the hurricane.

The decline in the population before Hurricane Maria may seem surprising but a closer inspection of the context yields two explanations for this phenomenon. 1) The dwindle in population coincides with the end of summer; hence, we expect tourists and visitors to leave the island around this time (20). Note that this is captured by both the Teralytics and APT data, which is expected. 2) The sharp decrease beginning a few weeks before the storm
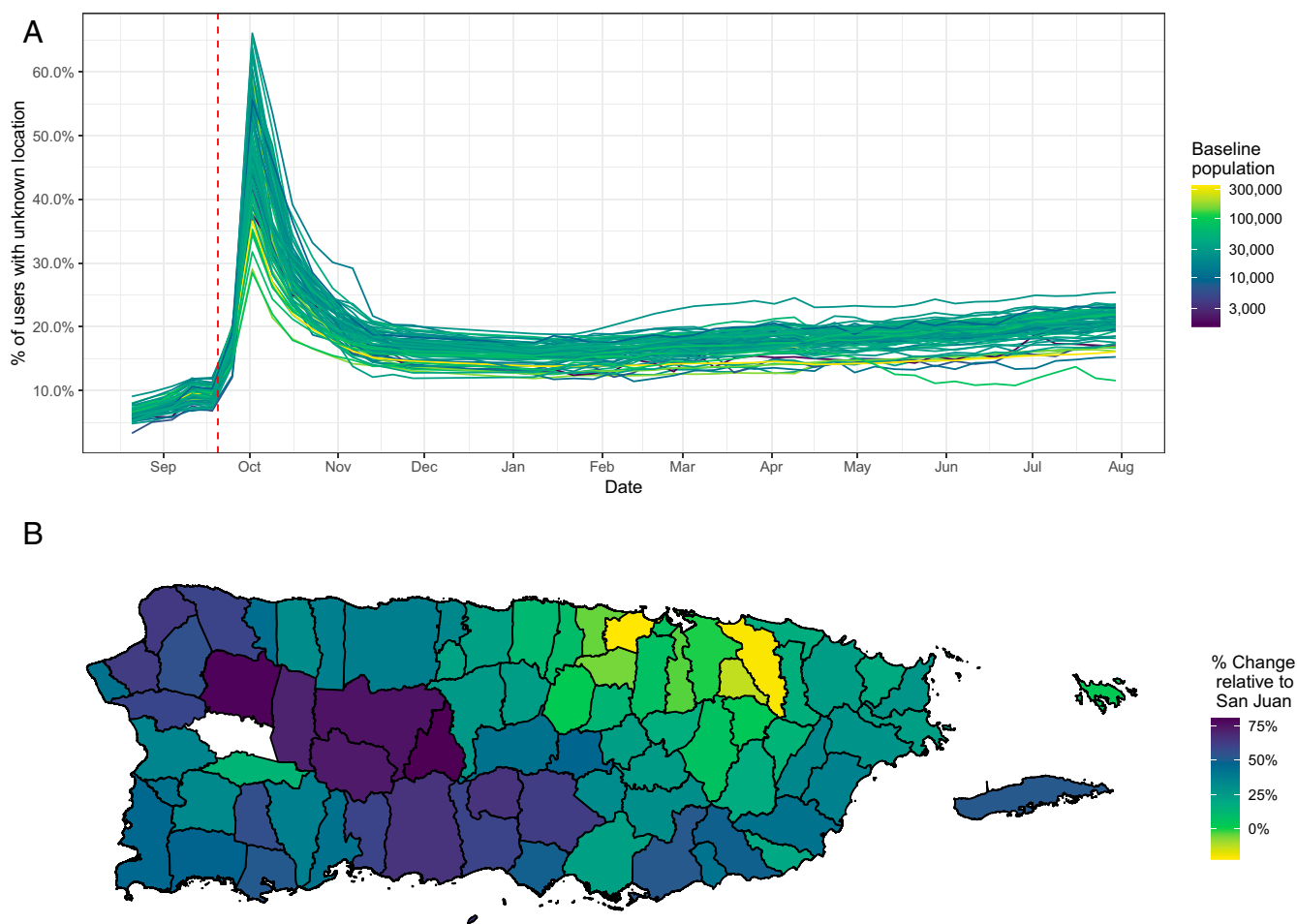
Fig. 4. Facebook users with unknown locations. (*A*) Percentage of individuals whose location was unknown by municipality. The vertical line denotes the day Hurricane Maria made landfall. Baseline population size is denoted in color. (*B*) Percentage of individuals with unknown location relative to San Juan for the week of 2 October 2017, the week after Hurricane Maria. The municipality in white represents Las Marias, the one municipality for which we had no data.

may be explained by the close encounter with Hurricane Irma, which scraped the northeastern part of the archipelago (21). Population displacement estimates based on the demographic balancing equation yield similar results (22). Although each has its own biases, these passively collected data estimates follow similar trends and show dynamic population fluctuations. In all cases, this decrease in population is markedly different from the trend since 2010 (*SI Appendix*, Fig. S4).

Our findings further suggest a rural–urban shift in population. Using Facebook data to analyze within-island movements, we observed persistent migration from rural to urban areas of Puerto Rico. This may be explained by individuals migrating from rural to urban areas in search of basic needs in the short term and staying due to increased access to resources in the long term. Out migration from rural areas continued in the Facebook cohort throughout the available data, becoming more concentrated in urban areas. Previous household surveys suggest greater out migration among younger individuals, potentially changing the demographic distribution of rural vs. urban regions in important ways (14).

In the immediate aftermath of a disaster, electricity, communication, and transportation are often affected, leading to sparse information on the areas most heavily affected. Assuming that individuals will continue to use services like mobile phones and Facebook if they had access, the lack of interaction with these services could serve as a proxy for damage from the hurricane. Spatial and temporal granularity in these immediately available datasets could augment satellite imagery and primary data sources

to more readily target priority areas for response. In the Facebook data, we can see that in the weeks immediately following Hurricane Maria, we identify areas that had larger proportions of cohort members whose locations were unknown compared with San Juan. Travel in most of these areas was impossible due to debris that was blocking streets and highways. Hence, this finding is not confounded with the rural–urban shift in the population described above. These areas coincide with rural municipalities found to be some of the first declared as disaster zones posthurricane (23).

Passively collected data provide a promising supplement to current at-risk population estimation procedures; however, each data source has its own biases and limitations. The population estimates from Facebook vs. Teralytics diverged significantly both before (by >9,500 people on 20 September 2017) and after (by >210,000 people on 20 April 2018) the hurricane. These data are not necessarily comparable since neither the market share of mobile phone providers nor Facebook users are likely to be perfectly representative of the overall population. In general, we expect those not included in the Facebook or Teralytics sample to be those without phones or internet access, which would likely include children, the elderly, and the very poor (24).

For privacy reasons, we were unable to analyze the demographic composition of the Facebook cohort, but these data represent a meaningful proportion of the Facebook population (25). Further, we showed that the percentage of users from each municipality is highly correlated with the baseline population size. The Facebook data also represent a closed cohort, unlike

the other sources. Membership into this group is defined during the 5-wk precrisis period, and no individuals can enter afterward. This means that this cohort will inherently only decrease in size. This rate of decrease is driven by 1) the baseline mortality rate, 2) the excess mortality due to the crisis, and 3) the baseline rate of discontinuation of Facebook use. One important advantage of Facebook data is the ability to analyze within-island movement patterns, which is not possible using flight or Teralytics estimates. Teralytics analyzed mobile phone data from a particular mobile operator of unknown market share and uses a proprietary algorithm to generate their population estimates, precluding any evaluation of primary sources or inherent biases. In both cases, it is impossible to accurately quantify the direction or magnitude of the biases. We have specified the benefits and drawbacks of each data source in *SI Appendix*, Table S1.

Despite the limitations of each dataset, passively collected data sources still provide a useful estimation of population displacement. The imputation procedure and the normalization of each dataset using baseline population values aim to ameliorate some of these limitations (*Methods*). The limitations of Facebook's data suggest that we are underestimating the true population size. However, the imputation procedure introduced here aims to mitigate that. Further, Facebook has expanded and introduced new datasets that do not suffer from the same limitations as the one used here (26). Research is needed to assess the capability of those datasets to estimate population size, with particular focus on expanding the study window to include data from previous or subsequent years and perhaps incorporate other similar regions during the same period for comparison. Given the paucity of data available immediately after a disaster, these data streams provide a clear benefit over the gold standard estimate for humanitarian purposes.

Our data show that emigration after disasters has a nonlinear effect on the count of population at risk and that this emigration is heterogeneous by rurality, affecting the denominators of many key population statistics. As interest in passively collected data grows and these tools are further refined to overcome current limitations, they can provide a more temporally and spatially nuanced picture of population movements after disasters.

Methods

ACS. We obtained yearly population estimates from 2010 to 2018 from the intercensal yearly population estimates provided by the ACS, a yearly survey conducted by the US Census Bureau. Specifically, we have the vintage 2018 estimates that were published last year.

Facebook. Facebook is a social media company with over 2 billion monthly active users. In 2017, the Data for Good team within Facebook launched Disaster Maps with the goal of aiding response organizations with information vital to optimal resource allocation in postdisaster settings. These maps are built using privacy-preserving techniques including aggregation and deidentification that protect individual privacy. The Harvard team accessed these data free of charge through Data for Good's standard license agreement, which allows partners working in humanitarian operations and research to improve their work through the use of Disaster Maps.

For this study, we used the original version of Facebook's Displacement Maps, a specific product within Disaster Maps (27), which has since been updated to improve its data sources and methodology. However, work is needed to assess the functionality of these improvements in the context described here. Displacement Maps are generated by first defining a geographic bounding box along with an index date defining the disaster event of interest. In this case, the geographic bounding box consisted of the entire island of Puerto Rico, and the index date was 20 September 2017, the day that Hurricane Maria made landfall on the island. In these data, Puerto Rico is divided into 188 nonoverlapping geographical tags (geotags), and users are assigned a home location defined as the geotags where they had the most interactions with the Facebook platform through a browser during the 5 wk prior to the index date. In the original Displacement Maps methodology, the location of these interactions was determined from associated internet protocol addresses; the new methodology instead utilizes location-based data from cell phones. Displacement Maps generate a closed population consisting of people using Facebook satisfying the following two conditions: 1) they registered an interaction with Facebook services from in the geographic

bounding box during the 5 wk preceding the index date, and 2) they were present in their home location during the week before the disaster (28).

They then followed this cohort through July 2018 and calculated the most commonly occurring geotag each week, aggregating total numbers of cohort members per geotag. Geotags with fewer than 100 people at baseline were excluded to protect privacy. Then, for each of the 49 wk after the index date, the dataset includes the number of crisis-affected people in their home location, the number of new users, and the number of unknown users for each geotag. Cohort members were defined as having an unknown location if they did not register an interaction during the week for which data were aggregated. For our results, we assumed these people with unknown locations were in their home location. We then combined geotag counts into Puerto Rico's 78 municipalities. Due to low counts, the municipality of Las María had no data. Below, we describe how we imputed other missing data.

Teralytics. Teralytics is a tech company that works with governments and private clients to assess human movement by partnering with mobile network operators. Specifically, the partnership allows the company to access and analyze the data that cell towers receive from mobile devices. From them, we obtained island-wide daily population proportion estimates relative to an undisclosed baseline from 31 May 2017 to 30 April 2018, based on all subscribers of a major undisclosed telecom company that created events in Puerto Rico. Events were defined as signal exchanges between a cell phone and the nearest cell phone tower. These signal exchanges occurred, for example, when a phone call was made or a text message was sent. The data were filtered by the provider, and only subscribers with activity all over the Teralytics analysis period were considered (31 May 31 2017 to 30 April 2018). Activity was defined as an event on at least 10 d/mo for all 12 mo. Due to the unreliability of data, proportion estimates for 4 wk after Hurricane Maria were not provided. For every day, a distinct number of subscribers in Puerto Rico is computed by considering events generated from different mobile devices. It is important to note that Teralytics is a commercial company that operates by competitively sourcing, cleaning, and preprocessing these data. Therefore, much of this analytic pipeline is proprietary and a black box to researchers. We have taken measures in our analysis to more readily index and compare this source with others; however, as noted in *SI Appendix*, Table S1, all of these sources have their benefits and drawbacks.

Aviation Records. Finally, through the Puerto Rico Institute of Statistics we obtained APT data from the US BTS. The data are composed of monthly counts of passengers who arrived and left the island per month from January 2010 to February 2018. The per-month difference between these two numbers will be referred to as net migration. We added the monthly net migrations to the vintage 2017 population estimates corresponding to the same date and interpolated using a linear model between these data points. This resulted in daily population estimates from July 2010 to July 2018 that account for flight passenger movement (29).

SI Appendix, Table S1 has a side-by-side comparison of all of the data sources. This research proposal was reviewed by the Harvard T. H. Chan School of Public Health Institutional Review Board and was deemed exempt as nonhuman subjects research.

Estimating Population Size. We estimated island-wide population sizes using each of the four data sources. For the ACS data, we simply interpolated the points for each year (*SI Appendix*, Fig. S3). For the other three sources, we defined population size estimate N_t for time t using the following:

$$N_t = N_0 \times \frac{m_t}{m_0},$$

where $N_0 = 3{,}325{,}001$, the ACS population estimate for 1 July 2017; m_t is a source-specific measurement for time t; and m_0 is a source-specific baseline. For the APT data, baseline was defined as 1 July 2017, m_t corresponds to the sum of N_0 and the cumulative net passenger movement for month t, and m_0 is the sum of N_0 and the cumulative net passenger movement at baseline. The formula above, therefore, is simply the cumulative sum, starting on 1 July 2017, of passengers arriving and passengers leaving:

$$m_0 = N_0 + (passengersout_0 - passengersin_0)$$

$$m_t = N_0 + \sum_{i=0}^{t}(passengersout_i - passengersin_i),$$

where i represents the ith month after baseline and $passengerout_i$ and $passengerin_i$ are the passengers leaving and entering Puerto Rico in month i, respectively. For the cell phone data, the baseline was also defined as 1 July

POPULATION BIOLOGY

2017; m_t and m_0 represent the proportions provided by Teralytics for time t and the proportion corresponding to baseline, respectively. Finally, for the Facebook data baseline was defined as 21 August 2017, which corresponds to the first observation in the dataset. Here, m_t corresponds to the Facebook population at time t, and m_0 represents the Facebook population at baseline.

For the municipality-level data, we aggregated the city-specific data from Displacement Maps by municipality in Puerto Rico while maintaining the temporal resolution at weeks. For all analyses, we assumed that cohort members who were defined as unknown for the week were present in their home town locations and were either unable to or chose not to interact with Facebook services using a browser in that period.

To evaluate the utility of Displacement Maps, a tool to target municipalities for resource allocation, we evaluated the proportion of the population with unknown locations every week compared with baseline. Our primary assumption here is that in the immediate aftermath of a disaster, any discontinuation of interaction with Facebook services defined at baseline would primarily be caused by loss of access to infrastructure, death, or other factors related to the event. Therefore, a higher proportion of municipality-specific cohort members whose location was unknown would be a proxy for impacts of the disaster in that region.

Imputation. In Facebook's dataset, Puerto Rico is divided into 188 nonoverlapping geotags. For each geotag, the dataset contains the number of crisis-affected people in their home location at time t, denoted here as home; the number of new users at time t, denoted here as new; and the number of users whose location is unknown at time t, denoted here as unknown. For privacy-preserving reasons, geotags were excluded from the data if they had fewer than 100 Facebook cohort members at time t. We, therefore, applied an imputation approach along with a sensitivity analysis to provide intervals of possible values. First, we computed the distribution of the maximum number of consecutive new missing values for each geotag. Thus, the jth geotag had a corresponding value, $c_j^{(new)}$, that represents the maximum number of consecutive weeks when the new variable was missing. Second, we denote a threshold τ that corresponds to the maximum number of successive weeks we are willing to accept and a default imputation value δ to

be used later. Then, we split the data into the geotags that comply with the threshold and those that did not [i.e., if $c_j^{(new)} \leq \tau$, then geotag j complies with the threshold]. For the geotags that complied with the threshold, we imputed the missing values using a linear interpolation of the observed values. For the geotags that did not comply with the threshold, we imputed the missing values with δ. Finally, we computed the distribution of the maximum number of consecutive home missing values for each tag. Hence, similar to above, the jth geotag has a corresponding value, $c_j^{(home)}$, that corresponds to the maximum number of consecutive weeks when the home variable was missing. Then, we excluded all geotags where the newly computed statistic was greater than τ (*SI Appendix*, Fig. S6). For our results, we used $\tau = 8$ and $\delta = 99$ (*SI Appendix*, Fig. S7). Note that $\delta = 99$ is the maximum number that the missing values can take. We conducted a sensitivity analysis where $\delta = 0$, the minimum number that the missing values can take (*SI Appendix*, Fig. S8). Our sensitivity analysis shows that the results do not change much.

1. L. Mbaye; African Development Bank Group, *Climate Change, Natural Disasters, and Migration* (IZA World Labor, 2017).
2. D. Thomaz, Post-disaster Haitian migration. *Forced Migration Rev.* **43**, 35–36 (2013).
3. K. J. Curtis, E. Fussell, J. DeWaard, Recovery migration after Hurricanes Katrina and Rita: Spatial concentration and intensification in the migration system. *Demography* **52**, 1269–1293 (2015).
4. The National Institute for Occupational Safety and Health (NIOSH), Methods: Mortality. CDC Append. https://wwwn.cdc.gov/eworld/Appendix/Mortality. Accessed 17 June 2019.
5. F. Checchi, L. Roberts, *Interpreting and Using Mortality Data in Humanitarian Emergencies. A Primer for Non-Epidemiologists* (Humanity Practice Networks, 2005), vol. 44.
6. OCHA, *The state of open humanitarian data: What data is available and missing across humanitarian crisies* (UN Office for the Coordination of Humanitarian Affairs, The Hague, The Netherlands, 2020).
7. United States Census Bureau, Methodology for the intercensal population and housing unit estimates: 2000 to 2010. https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/intercensal/2000-2010-intercensal-estimates-methodology.pdf. Accessed 11 June 2019.
8. R. Cruz-Cano, E. L. Mead, Excess deaths after Hurricane Maria in Puerto Rico. *J. Am. Med. Assoc.* **321**, 1005 (2019).
9. L. Bengtsson, X. Lu, A. Thorson, R. Garfield, J. von Schreeb, Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Med.* **8**, e1001083 (2011).
10. L. Bengtsson et al., Using mobile phone data to predict the spatial spread of cholera. *Sci. Rep.* **5**, 8923 (2015).
11. NOAA, Tropical cyclone report Hurricane Maria (AL152017). https://www.nhc.noaa.gov/data/tcr/AL152017_Maria.pdf. Accessed 15 May 2019.
12. C. D. Zorrilla, The view from Puerto Rico—Hurricane Maria and its aftermath. *N. Engl. J. Med.* **377**, 1801–1803 (2017).
13. C. Arnold, Death, statistics and a disaster zone: The struggle to count the dead after Hurricane Maria. *Nature* **566**, 22–25 (2019).
14. N. Kishore et al., Mortality in Puerto Rico after Hurricane Maria. *N. Engl. J. Med.* **379**, 162–170 (2018).
15. J. Sandberg et al., All over the place? Differences in and consistency of excess mortality estimates in Puerto Rico after Hurricane Maria. *Epidemiology* **30**, 549–552 (2019).
16. A. R. Santos-Lozada, J. T. Howard, Use of death counts from vital statistics to calculate excess deaths in Puerto Rico following Hurricane Maria. *J. Am. Med. Assoc.* **320**, 1491–1493 (2018).
17. R. Rivera, W. Rolke, Modeling excess deaths after a natural disaster with application to Hurricane Maria. *Stat. Med.* **38**, 4545–4554 (2019).
18. United States Census Bureau, More Puerto Ricans move to mainland United States, poverty declines. https://www.census.gov/library/stories/2019/09/puerto-rico-outmigration-increases-poverty-declines.html. Accessed 20 December 2019.
19. J. Schachter, A. Bruce, The impact of Hurricane Maria. https://unstats.un.org/unsd/demographic-social/meetings/2019/bangkok--intl-migration-workshop/Day3/Session9/5%20PR%20Hurricane%20Maria%20Impact%20presentation%20jps.pdf. Accessed 1 December 2020.
20. Foundation for Puerto Rico, Foundation for Puerto Rico's Visitor Economy Data Portal. https://www.foundationforpuertorico.org/visitoreconomydataportal. Accessed 3 August 2020.
21. J. P. Cangialosi, A. S. Latto, R. Berg, "Hurricane Irma 30 August–12 September 2017" (Tropical Cyclone Rep. AL112017, National Hurricane Center, Miami, FL, 2018).
22. A. R. Santos-Lozada, Revisiting the demography of disaster: Population estimates after Hurricane Maria. https://osf.io/preprints/socarxiv/n8vpe/ (28 March 2019).
23. R. Á. Claudio, 10 municipios de Puerto Rico declarados zona de desastre. Metro. https://www.metro.pr/pr/noticias/2017/09/12/10-municipios-de-puerto-rico-declarados-zona-de-desastre.html. Accessed 17 June 2019.
24. M. Duggan, J. Brenner, The Demographics of Social Media Users—2012, Pew Research Center (2013). https://www.pewresearch.org/internet/2013/02/14/the-demographics-of-social-media-users-2012/. Accessed 19 December 2019.
25. NapoleonCat, Facebook users in Puerto Rico–October 2018 (Facebook API). https://napoleoncat.com/stats/facebook-users-in-puerto_rico/2018/10. Accessed 19 December 2019.
26. Facebook Data for Good, Disaster Maps. https://dataforgood.fb.com/tools/disaster-maps/. Accessed 30 August 2020.
27. M. Jackman, Using data to help communities recover and rebuild. *Facebook Newsroom* (2017). https://about.fb.com/news/2017/06/using-data-to-help-communities-recover-and-rebuild/. Accessed 14 May 2019.
28. P. Maas et al., Facebook disaster maps: Methodology. *Facebook Research* (2017). https://research.fb.com/blog/2017/06/facebook-disaster-maps-methodology/. Accessed 14 May 2019.
29. J. Schachter, A. Bruce, Estimating Puerto Rico's population after Hurricane Maria: Revising methods to better reflect the impact of disaster. *The United States Census Bureau* (2020). https://www.census.gov/library/stories/2020/08/estimating-puerto-rico-population-after-hurricane-maria.html. Accessed 20 October 2020.
30. R. J. Acosta, N. Kishore, R. A. Irizarry, C. O. Buckee, Quantifying the dynamics of migration after Hurricane Maria in Puerto Rico. GitHub. https://github.com/RJNunez/pr-migration-paper. Deposited 1 December 2020.