















# Nowcasting Daily Population Displacement in Ukraine through Social Media Advertising Data

DOUGLAS R. LEASURE<sup>1</sup> , RIDHI KASHYAP<sup>2</sup> ,  
FRANCESCO RAMPAZZO<sup>3</sup> , CLAIRE A. DOOLEY<sup>4</sup> ,  
BENJAMIN ELBERS<sup>5</sup> , MAKSYM BONDARENKO<sup>6</sup> ,  
MARK VERHAGEN<sup>7</sup> , ARUN FREY<sup>7</sup> , JIANI YAN<sup>8</sup> ,  
EVELINA T. AKIMOVA<sup>9</sup> , MASOOMALI FATEHKIA<sup>10</sup> ,  
ROBERT TRIGWELL<sup>11</sup> , ANDREW J. TATEM<sup>12</sup> ,  
INGMAR WEBER<sup>13</sup>  AND MELINDA C. MILLS<sup>14</sup> 

*In times of crisis, real-time data mapping population displacements are invaluable for targeted humanitarian response. The Russian invasion of Ukraine on February 24, 2022, forcibly displaced millions of people from their homes including nearly 6 million refugees flowing across the border in just a few weeks, but information was scarce regarding displaced and vulnerable populations who remained inside Ukraine. We leveraged social media data from Facebook's advertising platform in combination with preconflict population data to build a real-time monitoring system to estimate subnational population sizes every day disaggregated by age and sex. Using this approach, we estimated that 5.3 million people had been internally displaced away from their baseline administrative region in the first three weeks after the start of the conflict. Results revealed four distinct displacement patterns: large-scale evacuations, refugee staging areas, internal areas of refuge, and irregular dynamics. While the use of social media provided one of the only quantitative estimates of internal displacement in the conflict setting in virtual real time, we conclude by acknowledging risks and challenges of these new data streams for the future.*

## Introduction

The Russian invasion of Ukraine on February 24, 2022, resulted in millions of refugees streaming into neighboring countries in the span of just a few weeks and millions more displaced from their homes inside Ukraine (UNHCR 2022). These vulnerable populations were in urgent need of targeted humanitarian assistance to ensure access to medical care, food,

water, shelter, and freedom of movement away from active conflict areas. While there was a swift and unified international response to the humanitarian crisis, particularly at the borders, there was a scarcity of information immediately following the invasion about the locations and demographics of displaced populations inside Ukraine. From a country of approximately 44 million people, border authorities documented nearly 6 million refugees crossing the border to flee the country in the first two months after the invasion leaving approximately 38 million people inside the country, many of whom were forced to abandon their homes or were trapped by ongoing conflicts (UNHCR 2022).

Rapid demographic estimation at the subnational level in the acute phase of a conflict is inherently challenging because preexisting population data quickly become outdated and new primary data collection may be extremely difficult or impossible (Robinson et al. 2003; Checchi et al. 2017; Ratnayake, Abdelmagid, and Dooley 2022). Qualitative methods based on key informants or other community-based information gathering (Henderson 2015) are often used to supplement displacement registry systems (Telford 1997) in the absence of surveys or other relevant data needed for quantitative population estimation approaches. Due to inaccessibility, quantitative methods that can be applied remotely are particularly useful. Satellite and aerial imagery have been used to map buildings and total settled area (Checchi et al. 2013) providing a basis for population estimation. With these approaches, however, field data from difficult-to-access areas are still needed to estimate current population densities (i.e., people per building or hectare of settled area) (Grais et al. 2006; Pinto et al. 2007), which may become quickly outdated as displacement trends change from day to day. Mobile phone location data have been proposed as one solution to monitor population flows without needing to collect field-based survey data (Bengtsson et al. 2011), but this approach is most useful when agreements are already in place with mobile phone operators to provide immediate access to data from crisis areas. Random digit dial telephone surveys have been used to conduct needs assessments for inaccessible areas in crises (Ashley and Scheuren 2010; Ruggiero et al. 2012), but this is not a standard approach for population estimation due to their often-limited sample sizes. For the weeks immediately following the Russian invasion of Ukraine, there was a moratorium on primary data collection inside the country and many areas were unsafe for field survey work which created a need for alternative sources of information to quantify rapidly changing population dynamics within the country.

Digital traces left behind when people use internet-based services provide an opportunity to fill these data gaps with anonymized and aggregated data on user locations in near real-time. A growing body of literature has leveraged digital traces for studying population dynamics, with a focus in particular on migration dynamics (Kashyap 2021; Rampazzo, Rango, and

Weber 2023). Studies on migration have used a diverse range of digital traces, including email IP addresses (Zagheni and Weber 2012), geolocated tweets (Zagheni et al. 2014), LinkedIn profiles (State et al. 2014), Skype calls (Kikas, Dumas, and Saabas 2015), among others (Cesare et al. 2018; Kashyap 2021). These novel sources of data have the advantage of being a byproduct of the use of digital technologies and platforms, and thus provide opportunities for real-time, dynamic measurement without requiring additional time and resources for primary data collection. However, they also come with their own unique challenges such as potentially hidden biases, algorithmic confounding, data access and licensing issues, and ethical considerations that are relevant to consider when repurposing these data for research (Kashyap 2021; Ratnayake, Abdelmagid, and Dooley 2022; Singh et al. 2021). Users of internet platforms are not a representative sample of the population and may exclude potentially vulnerable subpopulations who do not have mobile phones or internet access. While users of internet platforms like social media generally must consent to their data being used for marketing purposes, they do not explicitly provide consent for their data to be used for scientific studies or humanitarian responses. These data biases and ethical considerations must be carefully weighed when using these data and communicated to humanitarian organizations using population estimates derived from digital trace data for crisis response.

This study used a particular source of digital trace data—aggregate data from Facebook’s marketing tools, accessible via its marketing application programming interface (API)—which provides estimates of current audience sizes on Facebook for targeted advertising on the social media platform. Facebook’s marketing API provides counts of daily and monthly active users within specific age-sex demographic groups and subnational geographic areas. These data have previously been used for quantifying stocks of international migrants (Zagheni, Weber, and Gummadi 2017; Spyrtatos et al. 2019; Rampazzo et al. 2021) and to monitor flows of refugees in crisis situations such as the exodus of people from Venezuela in 2018–2019 (Palotti et al. 2020) and flows of outmigrants from Puerto Rico after Hurricane Maria in 2017 (Alexander, Polimis, and Zagheni 2019). Studies using these data to monitor stocks of migrants have generally either used nationally representative external data as training or validation data to support population-level inferences (e.g., Rampazzo et al. 2021). In other cases, such as in crisis settings, studies have either not attempted to extrapolate beyond the Facebook user population (e.g., Palotti et al. 2020) or used difference-in-difference approaches to monitor relative changes pre- and postcrises (e.g., Alexander, Polimis, and Zagheni 2019). Although population-level inferences are preferred for crisis response, representative survey data are often unavailable because preexisting data quickly become outdated and primary data collection is challenging if not impossible. Moreover, population changes in conflict and crises settings are highly dynamic, which require

high-frequency measurement that traditional data sources are often unable to provide.

To address these specific challenges of estimating subnational population displacement in a conflict setting, we sought to build a real-time monitoring system using data from Facebook's marketing API in combination with preconflict population data. Our goals were to estimate

- (1) Daily population sizes (i.e., including people not using Facebook) for age-sex demographic groups within subnational administrative units (Oblasts) of Ukraine,
- (2) Daily net changes in these populations relative to preconflict baseline population estimates, and
- (3) National total number of people internally displaced away from their baseline Oblast (level 1 administrative units) each day.

The data sources underlying our monitoring system included: (1) daily counts of active users on the Facebook social media platform (Meta 2022), (2) preconflict population estimates from the Common Operational Dataset for Population Statistics (COD-PS) of Ukraine (UNFPA 2022), and (3) daily counts of refugees crossing the border in and out of Ukraine each day (UNHCR 2022). Our approach was intended to detect inter-Oblast population displacement with the application of a demographic model assessing changing levels of Facebook penetration on a daily time step.

Our approach extends previous research using social media data for monitoring migration processes in three ways. First, in contrast to previous work using Facebook data in crisis contexts (e.g., Palotti et al. 2020; Alexander, Polimis, and Zagheni 2019), we provide population estimates at a daily frequency, highlighting how the integration of traditional and non-traditional data can meaningfully capture significant short-term changes. Second, we develop a simple demographic model to extrapolate beyond the Facebook population to estimate population changes in a setting where other sources of 'ground-truth' data are unavailable. Third, our method allows us to provide age- and sex-disaggregated subnational estimates, which highlight heterogeneity in demographic responses within conflict settings. In this way, our work makes a substantive contribution toward understanding the demographic consequences of the Russian invasion of Ukraine.

## Methods

The primary methodological challenge was to extrapolate age-sex population estimates from data on the Facebook user population. The urgent nature of the humanitarian crisis also posed challenges in being able to collect data and produce results immediately to inform the early stages of the humanitarian response when up-to-date demographic data were

**FIGURE 1** Proportional net changes in total population sizes for Ukrainian administrative units (Oblasts) as of June 21, 2022 (shown here as a proportion of preconflict baseline populations). Our study area included all Ukrainian Oblasts except the Autonomous Republic of Crimea and Sevastopol (Avtonomna Respublika Krym and Sevastopilska) because Facebook data were not available there.



unavailable to characterize the rapidly changing and increasingly vulnerable Ukrainian population.

### Study area

Our study area included all level-1 administrative units (Oblasts) inside Ukraine except those in the Crimea peninsula because no Facebook data were available there (Figure 1).

### Data

We utilized three data sources for demographic estimation: (1) preconflict baseline population estimates, (2) daily counts of border crossings leaving and entering Ukraine, and (3) counts of daily active users on Facebook. We present results alongside summaries of geolocated daily conflict events from ACLED (Raleigh et al. 2010) to provide relevant context.

*Baseline population.* We used baseline population estimates that were derived from the COD-PS corresponding to January 1, 2020 (UNFPA 2022).

This common operational dataset was used across United Nations agencies and other nongovernmental organizations for the humanitarian response in Ukraine. It provided subnational population estimates at the Oblast level (administrative level 1) (Kartographia Ukraine et al. 2022) disaggregated by age and sex. These estimates were population projections from the last population and housing census of Ukraine which was conducted in 2001.

These Oblast-level population projections from the COD-PS were disaggregated into 100 m grid cells with national coverage by Bondarenko et al. (2022) using a random forest machine learning approach that incorporated other geospatial covariates (Stevens et al. 2015; Bondarenko et al. 2021), most importantly the German Aerospace Center's remotely sensed impervious surfaces and World Settlement Footprint 3D building products from 2022 (Marconcini et al. 2021; Esch et al. 2022). The disaggregation method produced 100-m gridded population estimates that summed exactly to the original Oblast-level population totals from COD-PS. These gridded baseline population estimates provided flexibility for us to aggregate population estimates for geographic units used by the Facebook marketing API which differed slightly from the Oblast boundaries of the COD-PS. This step is robust to potential inaccuracies in the disaggregated population estimates at the 100-m spatial scale (i.e., the reaggregated population estimates for Facebook's geographic units were similar to the original COD-PS).

*Refugees.* We obtained daily counts of people crossing the border (i.e., leaving or entering Ukraine) from the United Nations High Commissioner for Refugees (UNHCR 2022). These data were collected by border authorities in each of the countries neighboring Ukraine. They did not include age-sex demographics of refugees, nor did they identify the source locations of refugees from inside Ukraine. With these data, for each day we subtracted the cumulative count of people entering Ukraine from the cumulative count of people leaving the country to obtain a cumulative net count of people leaving the country since the beginning of the conflict.

*Facebook users.* We collected counts of active Facebook users subnationally in Ukraine from the Facebook marketing API (Meta 2022) every day from February 25 through July 1, 2022, using *pySocialWatcher* software (Araujo et al. 2017; World Bank 2020). We collected data for users aged 13 and older disaggregated into five-year age classes for males and females separately. This included the following age classes: 13+, 18+, 18–60, 15–49, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, and 65+. These counts included all Facebook users whose “recent” location was in a given Oblast, as determined by their device data (Meta 2022). In other unpublished work outside Ukraine, we have seen lags of up to one week between internet blockages reported in the media and changes in Facebook's estimates of daily active users, so our results may also contain a

similar lag. Our data collections were set up to run continuously on a server with centralized database storage in a harmonized format.

Data from the Facebook marketing API can be accessed by anyone with a Facebook account and are intended for tracking audience sizes for targeted advertising campaigns on the social media platform. They are counts of active Facebook accounts (rather than users, per se) that may contain biases that arise from users with multiple accounts, fake accounts, and they contain some anomalous data such as sporadic zeros. It is unclear what data cleaning and model-based estimation that Facebook implements to generate these data. This results in some degree of quality uncertainty, random noise, and potentially unknown biases.

In Luhanska and Donetska Oblasts, user counts dropped to zero across all demographic groups on March 12, so we used data from March 11 from that point forward. In other areas where the threat of conflict resulted in decreasing Facebook users (e.g., Kyiv city and Kharkivska Oblast), we observed large reductions in Facebook users followed by a leveling off at a lower level. This provided some evidence suggesting that using the most recent dates with nonzero data was a reasonable approach in the absence of recent data for Luhanska and Donetska, which appeared to be leveling off at the time. Nonetheless, we acknowledge that this missed critical population fluctuations after this point in time in these Oblasts.

### Demographic estimation

We developed a deterministic model to estimate the daily population sizes by the age-sex demographic group in each Oblast based on counts of daily active Facebook users in each demographic group. This allowed us to estimate net changes in population for each Oblast every day beginning February 26 compared to the baseline on February 25.

The first step was to calculate the baseline Facebook penetration rate  $\psi_{i,s,a}$  for each Oblast  $i$ , sex  $s$ , age group  $a$ , and time  $t$ :

$$\psi_{i,s,a} = \frac{F_{i,s,a,t=0}}{N_{i,s,a,t=0}}, \quad (1)$$

where  $N_{i,s,a,t=0}$  is the baseline total population (i.e., 2020) and  $F_{i,s,a,t=0}$  is the baseline number of daily active users on Facebook from February 25, 2022.

We then calculated current population sizes as

$$N_{i,t,s,a} = \frac{F_{i,t,s,a}}{\psi_{i,s,a}}. \quad (2)$$

which assumed that the Facebook penetration rate had not changed since February 25, 2022 (i.e.,  $t = 0$ ). Because we expected that the penetration rates may have changed as the conflict continued, we anticipated that this could underestimate current populations. We applied a correction for this

below, but first we needed to account for age groups not represented on Facebook.

*65+ Populations.* The Facebook data combined all age groups 65 and over into a single age class, whereas the baseline population data included five-year age classes up to 80. We used the baseline age-sex proportions  $\rho_{i,s,a}$  to disaggregate current 65+ population estimates into five-year classes up to 80.

$$N_{i,t,s,a|a \geq 65} = \frac{\rho_{i,s,a} \sum_{a=65}^A N_{i,t,s,a}}{\sum_{a=65}^A \rho_{i,s,a}}. \quad (3)$$

Disaggregating the 65+ age groups in this way assumed that the current age-sex proportions in these age groups were the same as the baseline proportions.

*Under 20 populations.* The Facebook data contained no information for ages under 13 and had very few users less than 20 years old. To avoid these sparse data, we only used Facebook data for users 20+. We used baseline age-sex proportions to infer population sizes for age groups under 20 with the assumption that numbers of children were proportional to populations of reproductive age women (i.e., aged 20–49) rather than adult male populations or total adult populations.

To do this, we limited the following calculation to women of reproductive age (aged 20–49) and children (aged < 20). We first calculated the proportion of this subpopulation that was women of reproductive age  $\omega_i$  using baseline age-sex proportions  $\rho_{i,s,a}$ .

$$\omega_i = \frac{\sum_{a=20}^{49} \rho_{i,a,s=1}}{\sum_{s=1}^2 \sum_{a=0}^{19} \rho_{i,a,s} + \sum_{a=20}^{49} \rho_{i,a,s=1}}, \quad (4)$$

where  $s = 1$  refers to females and  $s = 2$  refers to males. We then combined this proportion with the population estimate for women of reproductive age derived from Facebook data (above) to estimate the daily population of children  $C_{i,t}$  in each Oblast.

$$C_{i,t} = \frac{\sum_{a=20}^{49} N_{i,t,a,s=1}}{\omega_i} - \sum_{a=20}^{49} N_{i,t,a,s=1}. \quad (5)$$

We disaggregated the total population of children into five-year age-sex classes using baseline age-sex proportions:

$$N_{i,t,s,a|a < 20} = \frac{C_{i,t} \rho_{i,s,a}}{\sum_{s=1}^2 \sum_{a=0}^{19} \rho_{i,s,a}}. \quad (6)$$

These age- and sex-specific population estimates for children assumed that population sizes in these age groups were proportional to the number of reproductive age women (i.e., aged 20–49) and that these proportions



were consistent with baseline proportions of children in the Oblast where populations were being estimated.

*Nonstationary Facebook penetration.* To account for potential changes in Facebook penetration rates through time, we imposed a constraint that required the regional age- and sex-specific population estimates to sum to the correct national total. However, this was complicated by the fact that millions of refugees had been flowing out of the country during this period. We helped control for this by constructing a scaling factor  $\theta_t$  that ensured the entire national population was accounted for in each time step. This used daily cumulative counts of refugees crossing the border to control for this by constructing a scaling factor that would ensure the entire national population was accounted for in each time step.

$$\theta_t = \frac{\left( \sum_{i=1}^I \sum_{s=1}^2 \sum_{a=1}^A N_{i,s,a,t} \right) - R_t}{\sum_{i=1}^I \sum_{s=1}^2 \sum_{a=1}^A N_{i,s,a,t}}, \quad (7)$$

where  $R_t$  is a cumulative net count of people crossing the border used to quantify refugee flows out of the country (UNHCR 2022). We then calculated an adjusted population estimate using this scaling factor.

$$\hat{N}_{i,t,s,a} = \theta_t N_{i,t,s,a}. \quad (8)$$

Applying the scaling factor ensured that our adjusted estimates of current populations  $\hat{N}_{i,s,a,t}$  in each Oblast and demographic group summed to the correct total nationally even if Facebook penetration rates had changed. This adjustment assumed that penetration rates changed evenly across regions and age-sex groups. We could not calculate more disaggregated scaling factors because we did not know the age-sex demographics of refugees crossing the border or their origin locations.

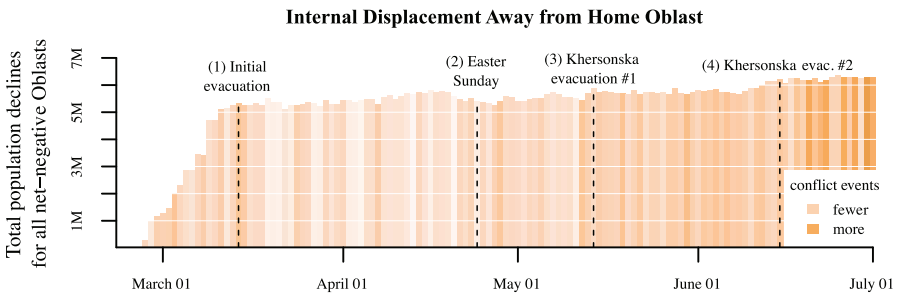
*Net population change.* We used the adjusted daily population estimates to calculate net changes in population  $\Delta_{i,s,a,t}$  relative to the baseline population from each Oblast:

$$\Delta_{i,s,a,t} = \hat{N}_{i,s,a,t} - N_{i,s,a,t} = 0. \quad (9)$$

This number was negative when net displacement was outward (i.e., declining populations) and positive when net displacement was flowing into an Oblast (i.e., increasing populations).

We created a metric of internal displacement by summing net population changes across all Oblasts where populations declined. While this metric did provide a valuable measure of internal displacement, it was not a comprehensive measure because it did not capture people displaced within an Oblast, and it could not detect net changes if numbers of people displaced away from an Oblast were offset by people being displaced into the same Oblast.

**FIGURE 2** Daily estimates of the number of people displaced away from their baseline Oblast. Shading indicates daily numbers of conflict events that were classified as “battles,” “explosions/remote violence,” or “violence against civilians” (Raleigh et al. 2010). Vertical dashed lines indicate a timeline of events: (1) March 14, initial evacuations particularly from Kyiv City, Kharkivska, and Kyivska Oblasts; (2) April 24, Orthodox Easter Sunday when nearly half a million people returned home; (3) May 14, the first evacuation of Khersonska Oblast; (4) June 15, the second evacuation of Khersonska Oblast.



## Results

Each day beginning February 26, 2022 (i.e., the first day we collected social media data), we estimated populations of males and females in five-year age classes from zero to 80+ years old for each Ukrainian Oblast, and from this we derived net changes in population sizes relative to preconflict baseline population estimates (UNFPA 2022) (see the Supporting Information). Net population changes at the Oblast level highlighted where populations were displaced from and to, and this estimate was updated on a daily basis (see Figure 1). Our study covered all Ukrainian Oblasts, except the Autonomous Republic of Crimea and Sevastopol because data from the Facebook marketing API were not available for these two Oblasts. Furthermore, although we were able to collect data for Donetska and Luhanska Oblasts, these were unreliable for two reasons: these regions had very low rates of Facebook usage even before the conflict, and user activity abruptly dropped to zero across all demographic groups on March 12, which we assumed was due to a loss of mobile phone, internet, and/or electrical infrastructure.

### National total internal displacement

We estimated the national total number of people displaced away from their baseline Oblast by summing net population changes across all Oblasts where populations had declined. While this will not capture all displacement between Oblasts, this approach is reasonable given that displacement into these net negative Oblasts is likely to be relatively low. Our measure of inter-Oblast population displacement increased sharply after the Russian invasion on February 24 reaching 5.3 million people by March 14 (Figure 2).

Inter-Oblast displacement fluctuated between 5 million and 6 million thereafter reaching a peak of 6.2 million people as of June 21.

As shown in Figure 2, this national-level metric of internal displacement derived using our real-time monitoring system was sensitive enough to detect key events on the ground, such as the evacuation of Khersonska Oblast immediately following a statement from its military-civilian administration on May 11 that it intended to bid for incorporation into Russia. Our Oblast-level population estimates suggested that more than 528,000 people (52 percent of the baseline population) left the Oblast by May 14, but that 522,000 of them had returned by May 17 likely due to widely reported road blockades by the Russian military preventing them from crossing into Ukrainian controlled areas of neighboring Oblasts. Consistent with media reports at the time, our national-level metric also captured a trend of more than 480,000 people returning to their home Oblasts during Orthodox Easter, with reductions of displaced people detected in 80 percent of Oblasts over the period from April 19 to April 27.

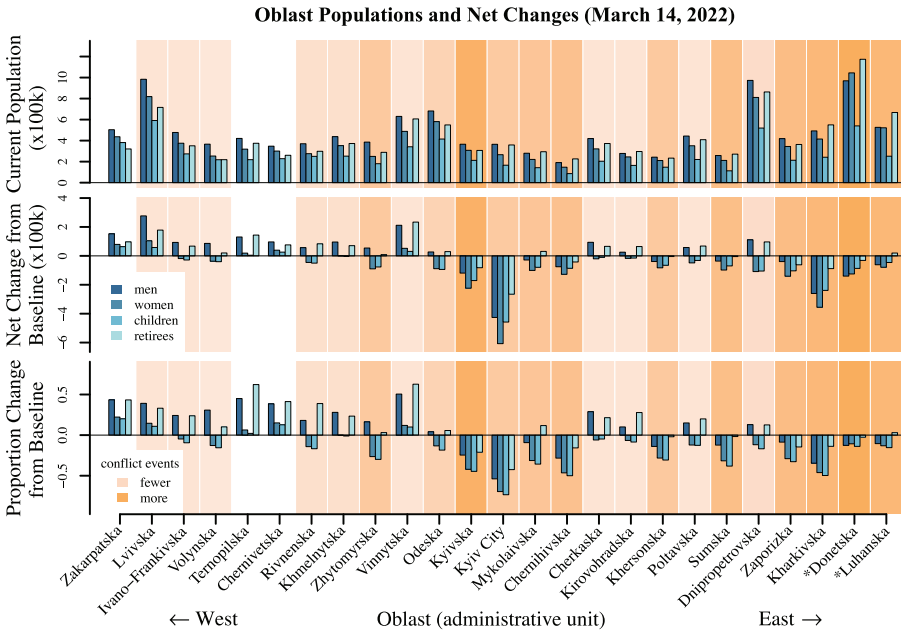
At the end of the initial evacuations between February 24 and March 14 (Figure 2), our results indicated that the population of Kyiv City had declined by nearly two-thirds, and the populations of Kharkivska, Kyivska, and Chernihivska Oblasts had declined by more than a third (Figure 3). Kyiv City and Kyivska Oblast together lost a staggering 2.3 million people during this 18-day period, and Kharkivska Oblast lost more than 943,000. Populations of women and children declined more than men and retirees (aged 60+) in these Oblasts, but all groups declined.

### Subnational displacement patterns

We provide estimates of populations and demographic changes for every Oblast (Figures 3 and 4). These results quantify large-scale evacuations of major cities in the first few weeks of the conflict, such as Kyiv and Kherson, and east-to-west movements of displaced persons during this period. They also show how geographic trends of displacement shifted through time with a large evacuation of Kherson coming months later. By June, populations of women (aged 20–59) and children (under 20) had decreased everywhere inside Ukraine except two western border Oblasts, whereas populations of men (aged 20–59) and retirees (both male and female aged 60+) increased in nearly all Oblasts with few or no conflict events. In the results that follow, we use these definitions of children (under 20), women (20–59), men (20–59), and retirees (60+) consistently. Results were similar for males and females over 60.

As highlighted in Figure 3, western Oblasts that experienced fewer conflict events saw a net population increase in the first three weeks of the conflict, whereas eastern Oblasts that experienced more conflict events saw a net population decrease. Increases in women and children were especially

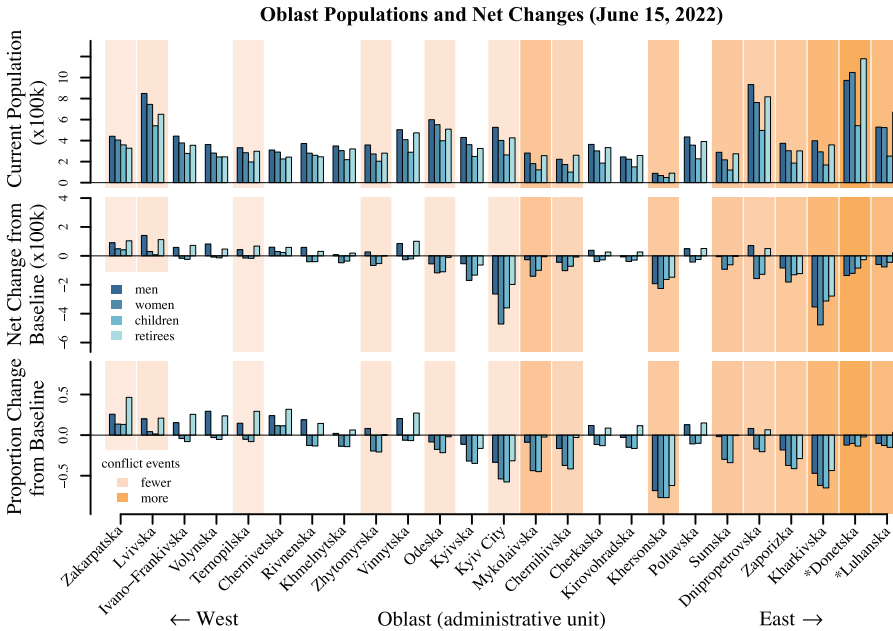
**FIGURE 3** Oblast populations and net changes from baseline as of March 14 for men and women (aged 20–59), children (aged 0–19), and retirees (aged 60+). Background shading indicates numbers of conflict events (log-scale) in the preceding 15 days that were classified as “battles,” “explosions/remote violence,” or “violence against civilians” (Raleigh et al. 2010). Oblasts are arranged from west to east. \*Results for Donetsk and Luhanska Oblasts were unreliable due to missing data.



notable in the western border Oblasts of Lvivska and Zakarpatska during the first three weeks of the conflict, and to a lesser extent in Chernivetska and Vinnytska. Women and children displaced to international border crossings were likely only there temporarily as they waited to flee the country. Our method of estimating populations of children (who are not active on Facebook) assumed their populations were proportional to women of reproductive age (see details in the Methods section).

Increases of men and retirees in these western border Oblasts were even more pronounced than for women and children, which may have been related to a moratorium on men aged 18–60 from leaving the country (Figure 3). We observed similar patterns of men and retirees accumulating in other western and central Oblasts where there had been relatively few conflict events. Populations of retirees declined the least of any demographic group in evacuated areas, and they appeared less likely than women and children to leave the country. Although retirees were not prohibited from leaving the country as men were, they may have been less able to embark on international journeys to seek refuge, opting instead to move shorter distances to relative safety in Oblasts with the least conflict.

**FIGURE 4** Oblast populations and net changes from baseline as of June 15 for men and women (aged 20–59), children (aged 0–20), and retirees (aged 60+). Background shading indicates numbers of conflict events (log-scale) in the preceding 15 days that were classified as “battles,” “explosions/remote violence,” or “violence against civilians” (Raleigh et al. 2010). Oblasts are arranged from west to east. \*Results for Donetsk and Luhanska Oblasts were unreliable due to missing data.

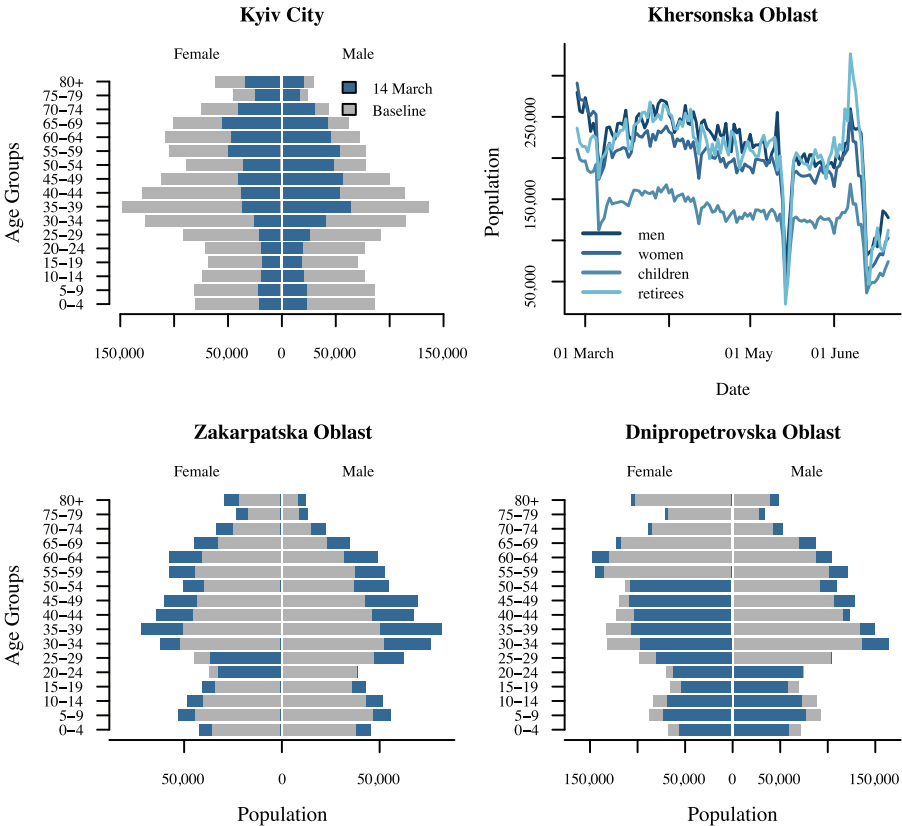


By June 15, there were only three Oblasts with an increase in the population of women and children compared to the baseline: Zakarpatska, Chernivetska, and to a lesser extent, Lvivska (Figure 4). Numbers of women and children had decreased in every other Oblast. In Oblasts with relatively few conflict events, a general trend of decreased women and children along with increases in men and retirees was still apparent.

At this point in time, the focus of the Russian military offensive had been redirected away from Kyiv with an increased focus on the east and south (Figure 4). Between March 14 and June 15, the population of Kyiv City rebounded from 1.2 million to 1.6 million, representing the gradual return of 26 percent of those who fled the city in the initial evacuation. During the same period, the population of Kharkivska Oblast declined from 1.7 million to 1.2 million, representing an additional 18 percent of the baseline population fleeing the region after 36 percent had already left during the initial evacuation.

As the Ukrainian military began an offensive to retake control of Khersonska Oblast in June, our results indicated that 583,000 people (57 percent of the baseline population) left the region between June 9 and June 15. Our

**FIGURE 5** Examples of demographic patterns associated with four types of displacement inside Ukraine. *Top-left*: Large-scale evacuations decreasing all demographic groups; *Bottom-left*: Refugee staging areas with increases across most demographic groups; *Bottom-right*: Internal safe havens for nonrefugees where only men and retirees increased; *Top-right*: Irregular population dynamics in the absence of open evacuation routes.



results suggested that this may have been a second attempt to leave the region for many of these people after half a million attempted to flee in mid-May only to return a few days later, potentially due to the Russian military preventing their movement into the Ukrainian-controlled territory.

### Demographic patterns of internal displacement

We identified four demographic patterns of internal displacement in Ukraine (Figure 5): (1) large-scale evacuations, (2) refugee staging areas, (3) internal safe havens for nonrefugees, and (4) irregular dynamics.

*Large-scale evacuations.* In the initial evacuations in early March, we observed reductions in all or most age-sex demographic groups resulting from large-scale evacuations. This occurred in Kyiv City, Kharkivska,

Kyivska, Donetska, Zaporizka, and Chernihivska Oblasts, and to a lesser extent in Sumska, Luhanska, Mykolaivska, and Khersonska Oblasts (Figures 3 and 4). In many of these locations, such as Kyiv City, populations of women declined more than men, but populations of all age-sex groups declined significantly.

*Refugee staging areas.* Oblasts with preferred international border crossings saw population increases across all demographic groups. We observed this pattern in early March for Zakarpatska, Lvivska, Chernivetska, and Vinnytska Oblasts, but the pattern persisted into June only for Zakarpatska and Chernivetska Oblasts (Figures 3 and 4). It appeared that women and children tended to transit through these regions while men and retirees were more likely to remain. This was potentially the result of a Ukrainian policy prohibiting most men aged 18–60 from leaving the country.

*Internal safe havens.* In Oblasts with relatively few conflict events but without preferred international border crossings, men and retirees tended to increase while women and children decreased or remained constant. This pattern was observed in Dnipropetrovska, Poltavska, Kirovohradska, Cherkaska, Khmelnytska, Rivnenska, Ternopilska, Volynska, and Ivano-Frankivska Oblasts (Figure 3). In Odeska Oblast, women and children left early in the conflict and there was a small increase in men and retirees (Figure 3), but by June there were reductions across all demographic groups likely due to changing perceptions of risks as the conflict evolved (Figure 4).

*Irregular dynamics.* Khersonska Oblast had irregular population dynamics that were similar across all demographic groups. There was an initial reduction in population estimates in early March concurrent with large-scale evacuations across the country, but numbers of men and retirees had rebounded in Khersonska Oblast by early April. There was a slow and steady decrease for all age groups throughout April followed by a large-scale evacuation after the May 11 announcement of intentions for Khersonska Oblast to join Russia. Most of these evacuees appeared to have returned within a few days potentially due to Russian military blockades preventing them from leaving the Oblast. At the beginning of June, there was an unexplained sharp increase in population estimates for all demographic groups followed by a second large-scale evacuation at about the same time as the Ukrainian military launched an offensive aiming to retake control of the region.

## Discussion

This work highlights the value of nontraditional data for nowcasting (i.e., estimate in near real-time) population dynamics at high frequency to support targeted humanitarian assistance in response to a crisis. We shared

results with the humanitarian community and the United Nations in the first few weeks following the Russian invasion and the daily frequency of digital trace data from the Facebook marketing API allowed us to produce daily population estimates through our real-time monitoring system. These were the only data-driven population estimates available at the time and continued to be the only estimates at the Oblast level with age-sex disaggregation for some time to follow, strengthening the capacity of the humanitarian response to target assistance for vulnerable groups such as children, elderly, and people trapped in active conflict areas. Our results highlighted different displacement patterns that varied geographically for women and children, elderly, and men, likely dependent on different risk assessments and migration options (Sánchez-Céspedes 2017; Curcio et al. 2019).

These results were used as one source of data triangulation to build confidence in estimates of internally displaced people (IDP) derived from other sources such as random digit dial telephone surveys conducted by the International Organization for Migration (IOM) in the early stages of the conflict (IOM 2022). The IOM random digit dial telephone surveys produced a representative sample of 2,000 people in each monthly survey round supporting statistical inferences about population displacement. However, this source of data collection could not achieve daily or weekly frequency to monitor rapidly changing population dynamics, nor did it have the statistical power to make inferences at the Oblast level, or provide demographic disaggregation by age and sex. In contrast, our real-time monitoring system relied on a combination of daily frequency social media advertising data spanning 4 million Facebook users with preconflict population estimates. Together, this combination of data enabled us to generate population displacement estimates at the finer geographical, temporal, and spatial resolution, thus complimenting other population datasets. The estimates derived from our approach were concurrent at the national level with those derived from the IOM telephone surveys. This converging evidence across two distinct, but complementary approaches, helped to build confidence in the UN's decision to significantly revise official estimates of IDPs from about 1.6 million to 6.48 million in mid-March (IOM 2022), bringing the scale of the internal displacement crisis in Ukraine into full focus.

Our work extended previous research that has used Facebook's marketing API for monitoring migration dynamics (e.g., Alexander, Polimis, and Zagheni 2019; Palotti et al. 2020; Rampazzo et al. 2021) by measuring subnational population displacement with daily frequency. Through automated, daily collections from Facebook's marketing API, we were able to build a real-time monitoring system to detect population changes in response to conflict events on the ground with the demographic disaggregation needed for humanitarian response and to support sampling designs for surveys of Ukrainians during the war (e.g., Wilson 2022). A central challenge in humanitarian response settings is that while population-level



inferences for subnational geographic units are required, representative validation data are not available. To partially overcome this challenge, and to enable us to generalize to population change beyond the Facebook user population, we implemented a data-driven constraint on the estimated total national population using available data on baseline populations and movements in and out of the country. Nevertheless, our approach relied on assumptions linking Facebook users to preconflict population baselines, and we acknowledge that these assumptions, as well as potentially unknown data quality issues, biases, and algorithmic confounding concerns that are inherent in social media advertising data (and digital trace data more broadly) also affect our study.

This work provides a proof-of-concept of how digital trace data can be leveraged in combination with population data to build real-time monitoring systems for rapid targeted assistance for vulnerable populations displaced by a crisis (e.g., armed conflict, natural disasters). Digital trace data provide several promises for this purpose through the finer temporal (e.g., daily) and geographical resolution they provide. However, there is much work that remains to be done to develop robust statistical methods to account for uncertainty and biases by combining multiple sources of information within such a monitoring system. As next steps, we see value in a Bayesian time series modeling approach to accommodate multiple sources of digital trace data and other data (e.g., from surveys) to inform a single process model. For example, because of inconsistencies across data types and concerns about double-counting individual users across multiple social media platforms, the deterministic model presented here was unable to incorporate data that were available from Instagram to better quantify younger populations and data that were available at the city-level from the V Kontakte social media platform which is more popular among Russian-speaking populations in eastern Ukraine where Facebook is used less. A Bayesian model would provide a potential framework to integrate across different social media platforms and data sources, which could help to develop robust estimates that are less sensitive to platform-specific fluctuations and outages. This could provide better population estimation for heavily impacted areas such as Luhanska and Donetsk Oblasts where our current results were unreliable.

A further area of extension is the integration of additional data sources, such as remotely sensed buildings (Esch et al. 2022) and georeferenced conflict locations (Raleigh et al. 2010). These geospatial data represent push-pull factors driving displacement patterns that could improve our daily population estimates and help further disaggregate our Oblast-level results into more fine-grained population estimates (e.g., 100-m grids). A higher degree of geographical resolution would allow the identification of specific locations potentially hosting large numbers of displaced people, as well as resident populations that remain in close proximity to active conflict events.

The humanitarian community would benefit from this detailed information for the targeted allocation of resources such as food provisions and health care services as well as for prioritizing additional data collection efforts (Ratnayake, Abdelmagid, and Dooley 2022). Building on machine learning methods designed to disaggregate population counts to high spatial resolution (Stevens et al. 2015), similar approaches have been developed to disaggregate counts of displaced populations based on the drivers of displacements patterns such as locations of conflict events or natural disasters, and desirable destinations such as relatively safe population centers and border crossings (Dooley et al. 2021). Combining these approaches with digital trace data is a very promising area for extension of this work. However, when doing so, it is critically important to balance potential benefits of targeted humanitarian assistance with potential harms that could result from mapping vulnerable populations in near real-time with a fine-grained spatial resolution (e.g., to prevent migration or otherwise target displaced populations; or for propaganda to claim external interference or dispute official numbers). In our case, we relied on a close partnership with our United Nations colleagues with relevant expertise and experience to help guide these decisions (OCHA Center for Humanitarian Data 2021) and also for future work we view partnering with organizations within the international humanitarian sector as a valuable mechanism for balancing the ethical and practical dimensions of this work.

## Conclusions

This work demonstrates the value of digital trace data to fundamentally shift the way in which we respond to global crisis events. We have shown how a real-time monitoring system, enabled by daily collections from the Facebook marketing API, facilitated daily nowcasts of population displacement to support the humanitarian community in the rapidly unfolding crisis situation following the Russian invasion of Ukraine. As we continue work to increase access and better understand these new types of data, we would urge that these approaches be seen as complementary to existing approaches rather than as replacements. It will always be prudent to triangulate multiple sources of information and methodologies when making rapid-response decisions with limited information in a crisis. The value of digital trace data is best realized in combination with other data sources, and, as in this study, we see significant value in the combination of digital trace with other forms of population data.

While digital trace data such as the social media advertising data used here provide significant opportunity, developing approaches to leverage these data in more sustainable and open ways in crisis response situations must address important challenges that confront access to these data. A key strength of the social media advertising data used in this study was their

accessibility through an API, which ensured more open modes of access. On the other hand, as these data are provided more directly for advertisers rather than for research or humanitarian response purposes, this created specific limitations in the features of the available data. For example, Facebook's marketing API provides no historic estimates, which made it essential but also computationally more demanding for our purpose to collect these data every day from the onset of the crisis. Streamlined and open data sharing and access through APIs, along with greater transparency about the provenance and data-generating processes of these data, are essential for ensuring their timely use in humanitarian settings.

There is a growing precedent of many companies holding large volumes of digital trace data opting for internal closed-door scientific use cases rather than through more general public release of data. Meta's Data for Good team have been particularly successful with this strategy, making valuable contributions that helped guide resources for refugees leaving Ukraine and to conduct a multicountry health survey in support of the global response to the COVID-19 pandemic (Salomon et al. 2021). Indeed (the job search engine) also published a compelling report on Ukrainian speakers using their platform to search for jobs in Poland, suggesting "urgent, open-ended job searches by people on the move" (Adrijan 2022). A global crisis, however, requires all hands on deck, and we cannot afford to exclude the vast expertise across the scientific, health, and humanitarian communities from accessing critical data in these situations. More open models of privacy-preserving data sharing between companies, researchers, and the humanitarian community, as well as legal frameworks that enable users to consent to data sharing with platforms explicitly for research purposes, are urgently needed to catalyze the effective and timely use of data for crisis situations. We encourage the research community, the humanitarian sector, and the technology companies who hold these powerful data to pursue open science together well in advance of the next crisis to provide transparency and confidence in new approaches that leverage these rich data for targeted humanitarian response.

## Acknowledgments

This data collection, analysis, and report were produced as a joint effort between the Leverhulme Centre for Demographic Science (LCDS; University of Oxford), Qatar Computing Research Institute (Hamad Bin Khalifa University), London School of Hygiene and Tropical Medicine, WorldPop (University of Southampton), and the United Nations International Organization for Migration. This project was approved by the Central University Research Ethics Committee at the University of Oxford (SSH\_SBS\_C1A\_21\_37). We thank Brian McDonald and Mohamed Bakr from the UN Migration Agency and Alessandro Sorichetta from WorldPop

for discussions in the first weeks of the humanitarian crisis. We want to acknowledge additional LCDS members and others for their support of this work including Per Block, Jen Comiskey, Jennifer Beam Dowd, Per Engzell, Ryan Gray, Clayton Leasure, Saul Newman, Andrea Tilstra, and Bo Zhao. This work was supported by the Leverhulme Trust under Grant RC-2018-003, the Bill and Melinda Gates Foundation under Grant INV-002697, the European Research Council under Advanced Grant 835079, and the Qatar Foundation.

## Supplementary Materials

Supplementary materials for this manuscript are available from an Open Science Foundation data repository (Leasure et al. 2022). The supplementary materials are described in detail in the online Appendix of this manuscript.

## Data and code availability

Meta's Platform Terms prohibit the authors from making the raw input data available for publication. However, comparable data can be accessed from the Facebook marketing API (Meta 2022; Araujo et al. 2017; World Bank 2020). All other data sources are cited, publicly available, and published with open licenses. Our full results with daily population estimates generated for five-year age-sex groups in each Oblast every day are available in the Supporting Information. The R code written for this analysis is also available in the Supporting Information.

## Consent to participate

It was not possible for the authors to request consent from Facebook users whose anonymized and aggregated data were used for this study. These users consented to the Meta Platform Terms which include consent for their data to be used for marketing purposes. We accessed counts of daily and monthly active Facebook users in the administrative level-1 region of Ukraine for five-year age-sex groups from Facebook's marketing tools which are available to anyone with a Facebook account.

---

## Notes

<sup>1</sup>Douglas R. Leasure, Leverhulme Centre for Demographic Science, University of Oxford, Oxford, UK; Department of Sociology, University of Oxford, Oxford, UK; Nuffield College, University of Oxford, Oxford, UK; WorldPop, School of Geogra-

phy and Environmental Science, University of Southampton, Southampton, UK. Email: douglas.leasure@demography.ox.ac.uk.

<sup>2</sup>Ridhi Kashyap, Leverhulme Centre for Demographic Science, University of Oxford, Oxford, UK; Department of Sociology,

University of Oxford, Oxford, UK; Nuffield College, University of Oxford, Oxford, UK.

<sup>3</sup>Francesco Rampazzo, Leverhulme Centre for Demographic Science, University of Oxford, Oxford, UK; Department of Sociology, University of Oxford, Oxford, UK; Nuffield College, University of Oxford, Oxford, UK.

<sup>4</sup>Claire A. Dooley, WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, UK; Department of Population Health, London School of Hygiene and Tropical Medicine, London, UK.

<sup>5</sup>Benjamin Elbers, Leverhulme Centre for Demographic Science, University of Oxford, Oxford, UK; Nuffield College, University of Oxford, Oxford, UK.

<sup>6</sup>Maksym Bondarenko, WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, UK.

<sup>7</sup>Mark Verhagen and Arun Frey, Leverhulme Centre for Demographic Science, University of Oxford, Oxford, UK; Department of Sociology, University of Oxford, Oxford, UK; Nuffield College, University of Oxford, Oxford, UK.

<sup>8</sup>Jiani Yan, Leverhulme Centre for Demographic Science, University of Oxford, Ox-

ford, UK; Department of Sociology, University of Oxford, Oxford, UK.

<sup>9</sup>Evelina T. Akimova, Leverhulme Centre for Demographic Science, University of Oxford, Oxford, UK; Department of Sociology, University of Oxford, Oxford, UK; Nuffield College, University of Oxford, Oxford, UK.

<sup>10</sup>Masoomali Fatehkia, Qatar Computing Research Institute, Hamad bin Khalifa University, Doha, Qatar.

<sup>11</sup>Robert Trigwell, International Organization for Migration, United Nations, Geneva, Switzerland.

<sup>12</sup>Andrew J. Tatem, WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, UK.

<sup>13</sup>Ingmar Weber, Qatar Computing Research Institute, Hamad bin Khalifa University, Doha, Qatar.

<sup>14</sup>Melinda C. Mills, Leverhulme Centre for Demographic Science, University of Oxford, Oxford, UK; Department of Sociology, University of Oxford, Oxford, UK; Nuffield College, University of Oxford, Oxford, UK; Department of Economics, Econometrics, & Finance, University of Groningen, Groningen, The Netherlands.

## References

- Adrjan, Pawal. 2022. "As War Rages, Ukrainian Language Job Searches Rise in Poland." <https://www.hiringlab.org/uk/blog/2022/06/02/ukrainian-language-job-searches-rise-in-poland/>
- Alexander, Monica, Kivan Polimis, and Emilio Zagheni. 2019. "The Impact of Hurricane Maria on Out-migration from Puerto Rico: Evidence from Facebook Data." *Population and Development Review* 45(3): 617–630. <https://doi.org/10.1111/padr.12289>.
- Araujo, Matheus, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. 2017. "Using Facebook Ads Audiences for Global Lifestyle Disease Surveillance: Promises and Limitations." In *WebSci '17: Proceedings of the 2017 ACM on Web Science Conference*. New York, NY: ACM. <https://github.com/joaopalotti/pySocialWatcher>
- Ashley, James D. and Fritz Scheuren. 2010. "Considerations in the Study Design of a Mobile Phone Survey of the Haitian Population." In *Joint Statistical Meetings Proceedings*, 5695–5702. Alexandria, VA: American Statistical Association.
- Bengtsson, Linus, Xin Lu, Anna Thorson, Richard Garfield, and Johan Von Schreeb. 2011. "Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti." *PLoS Medicine* 8 (8): e1001083. <https://doi.org/10.1371/journal.pmed.1001083>.
- Bondarenko, Maksym, D. Palacios-Lopez, Alessandro Sorichetta, Douglas R. Leasure, J. Zeidler, M. Marconcini, Thomas Esch, and Andrew J. Tatem. 2022. "Gridded Population Estimates for Ukraine Using UN COD-PS Estimates 2020, Version 2.0." WorldPop and German Aerospace Center (DLR), University of Southampton. <https://doi.org/10.5258/SOTON/WP00735>
- Bondarenko, Maksym, Jeremiah J. Nieves, Forrest R. Stevens, Andrea E. Gaughan, Warren Christopher Jochem, David Kerr, and Alessandro Sorichetta. 2021. "popRF: Random

- Forest-informed Population Disaggregation R package." Comprehensive R Archive Network (CRAN).
- Cesare, Nina, Hedwig Lee, Tyler McCormick, Emma Spiro, and Emilio Zagheni. 2018. "Promises and Pitfalls of Using Digital Traces for Demographic Research." *Demography* 55(5): 1979–1999. <https://doi.org/10.1007/s13524-018-0715-2>
- Checchi, Francesco, Barclay T. Stewart, Jennifer J. Palmer, and Chris Grundy. 2013. "Validity and Feasibility of a Satellite Imagery-Based Method for Rapid Estimation of Displaced Populations." *International Journal of Health Geographics* 12(1): 1–12. <https://doi.org/10.1186/1476-072X-12-4>
- Checchi, Francesco, Abdihamid Warsame, Victoria Treacy-Wong, Jonathan Polonsky, Mark Van Ommeren, and Claudine Prudhon. 2017. "Public Health Information in Crisis-Affected Populations: A Review of Methods and Their Use for Advocacy and Action." *The Lancet* 390(10109): 2297–2313. [https://doi.org/10.1016/S0140-6736\(17\)30702-X](https://doi.org/10.1016/S0140-6736(17)30702-X)
- Curcio, Carmen-Lucia, José H. Vanegas, María C. Palacio, and Jairo Corchuelo Ojeda. 2019. "Elderly and Forced Displacement in Colombia." *Colombia Medica* 50(2): 52–66.
- Dooley, Claire, Warren Jochem, Alessandro Sorichetta, Attila Lazar, and Andrew Tatem. 2021. "Description of Methods for South Sudan 2020 Gridded Population Estimates from Census Projections Adjusted for Displacement, Version 2.0." WorldPop, University of Southampton. <https://doi.org/10.5258/SOTON/WP00710>
- Esch, Thomas, Elisabeth Brzoska, Stefan Dech, Benjamin Leutner, Daniela Palacios-Lopez, Annkatrin Metz-Marconcini, Mattia Marconcini, Achim Roth, and Julian Zeidler. 2022. "World Settlement Footprint 3D-a First Three-Dimensional Survey of the Global Building Stock." *Remote Sensing of Environment* 270: 112877. <https://doi.org/10.1016/j.rse.2021.112877>
- Grais, Rebecca F., Denis Coulombier, Julia Ampuero, Marcelino E. S. Lucas, Avertino T. Barretto, Guy Jacquier, Francisco Diaz, Serge Balandine, Claude Mahoudeau, and Vincent Brown. 2006. "Are Rapid Population Estimates Accurate? A Field Trial of Two Different Assessment Methods." *Disasters* 30(3): 364–376. <https://doi.org/10.1111/j.0361-3666.2005.00326.x>
- Henderson, Alden. 2015. "Using the Delphi Method to Estimate Population Size and Demographics in Emergency Food Security Assessments." Emergency Food Security Assessments (EFSAs) Technical Guidance Sheet No. 10. Rome: World Food Programme. [https://documents.wfp.org/stellent/groups/public/documents/manual\\_guide\\_proced/wfp197300.pdf](https://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp197300.pdf)
- [IOM] International Organization for Migration. 2022. "Ukraine Internal Displacement Report: General Population Survey Round 1, 16 March 2022." IOM Ukraine, UN Migration. <https://displacement.iom.int/reports/ukraine-internal-displacement-report-general-population-survey-round-2-9-16-march-2022>
- Kartographia Ukraine, United Nations Office for the Coordination of Humanitarian Affairs, and Information Technology Outreach Service. 2022. "Ukraine Subnational Administrative Boundaries. United Nations Office for the Coordination of Humanitarian Affairs." <https://data.humdata.org/dataset/cod-ab-ukr>
- Kashyap, Ridhi. 2021. "Has Demography Witnessed a Data Revolution? Promises and Pitfalls of a Changing Data Ecosystem." *Population Studies* 75(Supp. 1): 47–75. <https://doi.org/10.1080/00324728.2021.1969031>
- Kikas, Riivo, Marlon Dumas, and Ando Saabas. 2015. "Explaining International Migration in the Skype Network: The Role of Social Network Features." In *Proceedings of the 1st ACM Workshop on Social Media World Sensors*, 17–22. New York: ACM.
- Leasure, Douglas R., Ridhi Kashyap, Francesco Rampazzo, Benjamin Elbers, Claire A. Dooley, Ingmar Weber, Masoomali Fatehikia., et al. 2022. "Nowcasting daily population displacement in Ukraine through social media advertising data: Supplemental Materials." Leverhulme Centre for Demographic Science, University of Oxford; SocArXiv. <https://doi.org/10.31235/osf.io/6j9wq>
- Marconcini, Mattia, Annkatrin Metz-Marconcini, Thomas Esch, and Noel Gorelick. 2021. "Understanding Current Trends in Global Urbanization: The World Settlement Footprint Suite." *12th International Symposium on Digital Earth* 9(1): 33–38. [https://doi.org/10.1553/giscience2021\\_01\\_s33](https://doi.org/10.1553/giscience2021_01_s33)
- Meta. 2022. "Marketing API, v13.0." Meta. <https://developers.facebook.com/docs/marketing-apis>

- OCHA Center for Humanitarian Data. 2021. "OCHA Data Responsibility Guidelines." United Nations Office for Coordination of Humanitarian Affairs. [https://data.humdata.org/dataset/2048a9475714-4220-905b-e662cbcd14c8/resource/60050608-0095-4c11-86cd0a1fc5c29fd9/download/ocha-data-responsibility-guidelines\\_2021.pdf](https://data.humdata.org/dataset/2048a9475714-4220-905b-e662cbcd14c8/resource/60050608-0095-4c11-86cd0a1fc5c29fd9/download/ocha-data-responsibility-guidelines_2021.pdf)
- Palotti, Joao, Natalia Adler, Alfredo Morales-Guzman, Jeffrey Villaveces, Vedran Sekara, Manuel Garcia Herranz, Musa Al-Asad, and Ingmar Weber. 2020. "Monitoring of the Venezuelan Exodus Through Facebook's Advertising Platform." *PLoS ONE* 15(2): e0229175. <https://doi.org/10.1371/journal.pone.0229175>
- Pinto, A., V. Brown, K. Chan, I. Chavez, S. Chupraphawan, R. Grais, P. Lai, S. Mak, J. Rigby, and P. Singhasivanon. 2007. "Estimating Population Size Using Spatial Analysis Methods." In *GIS for Health and the Environment*, 271–287. Berlin: Springer. [https://doi.org/10.1007/978-3-540-71318-0\\_20](https://doi.org/10.1007/978-3-540-71318-0_20).
- Raleigh, Clionadh, Rew Linke, Håvard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED Armed Conflict Location and Event Data." *Journal of Peace Research* 47(5): 651–660. <https://doi.org/10.1177/0022343310378914>.
- Rampazzo, Francesco, Jakub Bijak, Agnese Vitali, Ingmar Weber, and Emilio Zagheni. 2021. "A Framework for Estimating Migrant Stocks Using Digital Traces and Survey Data: An Application in the United Kingdom." *Demography* 58(6): 2193–2218. <https://doi.org/10.1215/00703370-9578562>.
- Rampazzo, Francesco, Marzia Rango, and Ingmar Weber. 2023. "New migration data: Challenges and Opportunities." In *Handbook of Computational Social Science for Policy* edited by Eleonora Bertoni, Matteo Fontana, Lorenzo Gabrielli, Serena Signorelli, and Michele Vespe, pp. 345–359. Cham: Springer. [https://doi.org/10.1007/978-3-031-16624-2\\_18](https://doi.org/10.1007/978-3-031-16624-2_18)
- Ratnayake, Ruwan, Nada Abdelmagid, and Claire Dooley. 2022. "What We do Know (and could know) about Estimating Population Sizes of Internally Displaced People." *Journal of Migration and Health*, 6: 100120. <https://doi.org/10.1016/j.jmh.2022.100120>
- Robinson, Courtland, Elizabeth Rowley, Paul Bolton, William Weiss, and Kavita Singh. 2003. *Demographic Methods in Emergency Assessment: A Guide for Practitioners*. Baltimore, Maryland, USA: Centre for International Emergency, Disasters and Refugee Studies, Johns Hopkins University Bloomberg School of Public Health.
- Ruggiero, Kenneth J., Heidi S. Resnick, Lisa A. Paul, Kirstin Gros, Jenna L. McCauley, Ron Acierno, Mark Morgan, and Sandro Galea. 2012. "Randomized Controlled Trial of an Internet-Based Intervention Using Random-Digit-Dial Recruitment: The Disaster Recovery Web Project." *Contemporary Clinical Trials* 33(1): 237–246. <https://doi.org/10.1016/j.cct.2011.10.001>.
- Salomon, Joshua A., Alex Reinhart, Alyssa Bilinski, Eu J. Chua, Wichida La Motte-Kerr, Minttu M. Rönn, Marissa B. Reitsma., et al. 2021. "The US COVID-19 Trends and Impact Survey: Continuous Real-Time Measurement of COVID-19 Symptoms, Risks, Protective Behaviors, Testing, and Vaccination." *Proceedings of the National Academy of Sciences* 118(51): e2111454118. <https://doi.org/10.1073/pnas.2111454118>
- Sánchez-Céspedes, Lina M. 2017. "The Consequences of Armed Conflict on Household Composition." *Oxford Development Studies* 45(3): 276–302. <https://doi.org/10.1080/13600818.2016.1213798>
- Singh, Neha S., Michelle Lokot, Chi-Chi Undie, Monica A. Onyango, Rosemary Morgan, Anne Harmer, Jane Freedman, and Shirin Heidari. 2021. "Research in Forced Displacement: Guidance for a Feminist and Decolonial Approach." *The Lancet* 397(10274): 560–562. [https://doi.org/10.1016/S0140-6736\(21\)00024-6](https://doi.org/10.1016/S0140-6736(21)00024-6)
- Spyratos, Spyridon, Michele Vespe, Fabrizio Natale, Ingmar Weber, Emilio Zagheni, and Marzia Rango. 2019. "Quantifying International Human Mobility Patterns using Facebook Network Data." *PLoS ONE* 14(10): e0224134. <https://doi.org/10.1371/journal.pone.0224134>
- State, Bogdan, Mario Rodriguez, Dirk Helbing, and Emilio Zagheni. 2014. "Migration of Professionals to the US." In *Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11–13, 2014. Proceedings*, 531–543. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-13734-6\\_37](https://doi.org/10.1007/978-3-319-13734-6_37)

- Stevens, Forrest R., Andrea E. Gaughan, Catherine Linard, and Andrew J. Tatem. 2015. "Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data." *PLoS ONE* 10(2): e0107042. <https://doi.org/10.1371/journal.pone.0107042>.
- Telford, John. 1997. *Counting and Identification of Beneficiary Populations in Emergency Operations: Registration and Its Alternatives*. London, UK: Overseas Development Institute (ODI).
- UNFPA. 2022. "Ukraine Common Operational Dataset on Population Statistics (COD-PS)." United Nations Office for the Coordination of Humanitarian Affairs. <https://data.humdata.org/dataset/cod-ps-ukr>
- UNHCR. 2022. "Refugees Fleeing Ukraine." United Nations High Commissioner for Refugees. <https://data2.unhcr.org/en/situations/ukraine>
- Wilson, Alexandra. 2022. "How do Ukrainians See the War Playing Out?" Premise. <https://www.premise.com/blog/how-do-ukrainians-see-the-war-playing-out/>
- World Bank. 2020. "Connectivity Maps Using Advertisement Platforms." The World Bank. [https://worldbank.github.io/connectivity\\_mapping/intro.html](https://worldbank.github.io/connectivity_mapping/intro.html)
- Zaghenni, Emilio, Venkata R. K. Garimella, Ingmar Weber, and Bogdan State. 2014. "Inferring International and Internal Migration Patterns from Twitter Data." In *Proceedings of the 23rd International Conference on World Wide Web*, 439–444. <https://doi.org/10.1145/2567948.2576930>
- Zaghenni, Emilio, and Ingmar Weber. 2012. "You are Where You e-mail: Using e-mail Data to Estimate International Migration Rates." In *Proceedings of the 4th Annual ACM Web Science Conference*, 348–351. New York: ACM. <https://doi.org/10.1145/2380718.2380764>
- Zaghenni, Emilio, Ingmar Weber, and Krishna Gummadi. 2017. "Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants." *Population and Development Review* 43(4): 721–734. <http://www.jstor.org/stable/26622775>