Journal of Migration and Health 6 (2022) 100120

Contents lists available at ScienceDirect

Journal of Migration and Health

journal homepage: www.elsevier.com/locate/jmh

What we do know (and could know) about estimating population sizes of internally displaced people

Ruwan Ratnayake[a,b,c,*], Nada Abdelmagid[a,b], Claire Dooley[b,d]

[a] Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom
[b] Health in Humanitarian Crises Centre, London School of Hygiene and Tropical Medicine, London, United Kingdom
[c] Centre for the Mathematical Modeling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom
[d] Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

ARTICLE INFO

Keywords:
Humanitarian assistance
Internal displacement
Population

ABSTRACT

The estimation of population denominators of internally displaced people (IDP) and other crisis-affected populations is a foundational step that facilitates all humanitarian assistance. However, the humanitarian system remains somewhat tolerant of irregular and inaccurate estimates of population size and composition, particularly of IDPs. In this commentary, we review how humanitarian organizations currently approach the estimation of IDP populations, and how field approaches and analytical methodologies can be improved and integrated.

The estimation of population denominators of crisis-affected populations is a foundational step that facilitates all humanitarian assistance. It is needed to assess the scale of the crisis and its resource needs, monitor the population's health status, access to services, and risk of epidemics, and if disaggregated data is available, to identify subgroups at increased risk (Sphere 2018; Abdelmagid and Checchi, 2018). The term 'crisis-affected' refers to different types of populations, including residents, internally displaced persons (IDP), refugees, and returnees (Diggle et al., 2017).

However, the humanitarian system remains somewhat tolerant of irregular and inaccurate estimates of population size and composition (Diggle et al., 2017; Checchi et al., 2017; André, 2020). Despite using standard approaches, humanitarian efforts to estimate IDP populations during recent large-scale displacements in Burkina Faso, Cameroon, Syria, and Yemen are likely affected by inaccuracies in population estimation (Diggle et al., 2017; André, 2020; Checchi and Koum Besson, 2022). The impacts of this underestimation are varied and include the over- or under-estimation of population denominators and therefore inaccuracy in morbidity, mortality, and service coverage indicators that are used to facilitate vaccination campaigns, health service planning, and overall resource allocation (Diggle et al., 2017; Bowden et al., 2012). Poor humanitarian access to populations may prevent efforts to collect population data (Diggle et al., 2017). Persistent insecurity may trigger repeated movements by IDP communities, which become difficult to measure (André, 2020). Needs assessments that are based on surveying these populations and their demographic groups may be biased by the inaccurate initial estimates. By the end of 2020, the United Nations High Commissioner for Refugees (UNHCR) estimated that 82.4 million persons were forcibly displaced globally, 48 million of whom were IDPs (UNHCR 2021). This latter number falls somewhere between the populations of Spain and Kenya (Total population). This raises the question: how do governments and humanitarian organizations approach the estimation of IDPs, and how can the field approaches and analytical methodologies be improved?

Estimating IDP populations in the acute phase of a crisis is particularly challenging given that estimation needs to be equally dynamic and systematic to capture populations in flux (Abdelmagid and Checchi, 2018; Checchi et al., 2017). Enumeration is done in a chaotic context, where insecurity can render data collection efforts risky and humanitarian access can be severely restricted, as recently seen in Northern Nigeria and Anglophone areas of Cameroon (André, 2020). IDPs may be displaced many times and may return to their place of origin, all within a short period. Further along the timeline of displacement, IDPs increasingly take up residence in urban and peri-urban areas outside of camp settings where there may be more economic opportunities and a larger network of persons from their own communities (Spiegel et al., 2010). A cogent timeline for displacement may also be difficult to establish as different families may move according to different timelines, away from

* Corresponding author at: Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom.
E-mail address: ruwan.ratnayake@lshtm.ac.uk (R. Ratnayake).

https://doi.org/10.1016/j.jmh.2022.100120
Received 28 February 2022; Received in revised form 21 May 2022; Accepted 28 May 2022
Available online 29 May 2022

the larger group. Similar to refugees fearing refoulement, IDPs may be concerned about retribution from host communities, and therefore willing to remain hidden, largely cutoff from humanitarian assistance, and effectively uncounted.

Population estimation by actual enumeration of persons or structures is typically based on three field methods: an initial door-to-door census followed by population-based demographic surveillance (as the gold standard); field surveys of residential structures and estimation by mean household size; measurement of the spatial area and estimation by population density; and convenience-based methods, including the analysis of routine vaccination data, estimation based on age structure, and the use of key informants to estimate population sizes (Abdelmagid and Checchi, 2018; Checchi et al., 2017). Crude data is sometimes, but not always, disaggregated to reveal the basic demographic profile by age and sex. This helps to identify the relative burden of groups at risk, including children less than five years of age, women of reproductive age and elderly persons. Building on these field methods are remotely-applied approaches which incorporate statistical modeling of satellite imagery of residential structures or population density from remotely-sensed climate, illumination, and other land-use variables (Checchi and Koum Besson, 2022). Estimation of IDP population flow from one place to another and forecasting of IDP population sizes and destinations can also be done (Checchi and Koum Besson, 2022; Huynh and Basu, 2020). These efforts are reliant on the availability of repeated satellite images, expertise in statistics and modeling, and the benefit of the time needed to rigorously execute these analyses (Global Health Cluster 2018; Checchi et al., 2013). Critically, the quality of the forecasting efforts is reliant on the quality, completeness, and representativeness of the initial field data that is collected. This links data collection efforts directly to forecasting. And so, the time may be right for further research and validation of promising field-based, remote-enhanced, and combined field and remote approaches that show potential for estimating IDP populations in a more dynamic and resource-saving manner (Franklinos et al., 2021; Ratnayake et al., 2020).

IDP population estimation will always be based on primary data collected from the population itself. Therefore, its major challenge is providing the logistical resources and humanitarian access to carry out repeated estimation at monthly or near-monthly frequency to capture changing displacement scenarios (Checchi et al., 2017). In the acute phase however, this means that estimates derived with traditional methods are quickly outdated by continuous population flux. This is typified by recent scenarios of large-scale forced migration to informal settlements in Afghanistan, around Bangui in Central African Republic, and southern Yemen. Rapid estimation in these contexts requires addressing both the dynamic nature of movement and conducting this estimation at a large enough geographical scale to be comprehensive (i.e., across several informal settlements). Where IDPs make up the majority of households, residential structure counting based on satellite imagery has been automated using machine learning methods with serial, very high resolution (VHR) satellite images for structures in diverse locales including in Bangladesh, Jordan, Somalia, South Sudan, and Yemen (Global Health Cluster, 2018; Koum Besson et al., 2021; Quinn et al., 2018). Similarly, remote sensing based on modeling of population density build-up can be done efficiently based on freely-available (albeit annually-derived) maps from the WorldPop project (Tatem, 2017). These methods provide an estimate of the entire contiguous population that can be sensed through the map, unless VHR satellite imagery shows specific IDP camps, slums or informal settlements and the user knows the specific locations where IDPs have settled and timely information on the displacement events which have triggered new movements. Currently, the International Organization for Migration (IOM) operates the Displacement Tracking Matrix (DTM) program to track such events through key informants (International Organization for Migration IOM, 2021). On the other hand, when equating IDP residence with actual visible residential structures, undercounting of persons in highly-dense residences, multi-story dwellings, dilapidated and temporary housing structures and of families that lack shelter remains an epidemiological concern.

The use of call detail records to track movements of IDPs has also been attempted in several urban and peri-urban humanitarian contexts where it is likely that IDPs own working mobile phones (Bharti et al., 2015). Since its validation during the 2010 earthquake in Haiti (Bengtsson et al., 2011), it has been utilized to estimate IDP populations and IDP flows between locations in crises where data sharing agreements with network operators had been setup prior, including in Côte d'Ivoire, Mozambique, and Nepal (Li et al., 2019). The availability of real time information on vulnerable populations to warring parties, and the underestimating population sizes based on inequitable mobile phone ownership by IDPs remain protection and epidemiological concerns, respectively (Princeton School of Public and International Affairs, 2021). The proportion of IDP families with at least one working mobile phone is likely to vary greatly depending on the geographical setting and the level of penetration of mobile phones in that area. All told, estimation based on the known proportion of IDPs with mobile phones could be a promising means of collecting routine population data.

Key challenges to these remote methods using satellite imagery or call detail records include the procurement of timely and inexpensive serial satellite images of local areas (i.e., towns and cities) that are sufficiently high-resolution to visualize residential structures and in developing agreements for call detail records with network operators. One promising development are open datasets such as Open Buildings, which aims to map building footprints and may be freely-available (https://sites.research.google/open-buildings/) or obtained at cost. Similar mapping of building footprints has been used to estimate populations in 100 m grid cells in the Democratic Republic of the Congo (DRC) (Boo et al., 2022; Ecopia, 2022). In some settings like post-Hurricane Maria in Puerto Rico, Venezuela, and currently in Ukraine, Facebook's Disaster Maps project and/or its marketing data has been used to provide aggregated and deidentified data on the scale and movement of users within the affected areas (Maas et al., 2019; Acosta et al., 2020; Palotti et al., 2020; Leasure et al., 2022). Building on this, there is a central need to proactively combine accurate, and systematic field-based estimation efforts with automation, that is the judicious use of machine learning methods for repetitive counting tasks. As with any statistical estimation method, there is also a need to incorporate measures of statistical uncertainty, and to posit relatively narrow lower and upper bounds for population sizes, instead of a single static estimate, or widely-bounded estimates (that are therefore less operationally usable).

An ensuing but no less important challenge is for the estimation of an effectively hidden IDP population in non-contiguous areas. This is a common scenario wherein IDPs are nested within host populations, as has been seen in Syria, South Sudan, and the DRC (Spiegel et al., 2010). Community-based demographic surveillance is often used to map and estimate IDP populations in small and medium sized communities (Ratnayake et al., 2020; Baaees et al., 2021). In a population-based household survey among hidden Syrian refugee households in urban Jordan, a local network of community health workers (CHW) was used to facilitate refugees to self-identify their household and their neighbors for the purpose of a refugee-specific health assessment (Ratnayake et al., 2020). Here, a geographic information system was used to overlay a grid of the city and probabilistic sampling of grid cells was used to provide a starting point for the Syrian CHWs to create a snowball sampling network of refugee households in each grid cell. In certain contexts where communities deem this interaction acceptable, this method could be extended to IDPs to self-identify local networks of households exhaustively and rapidly. It is conceivable that estimates from CHWs could also be routinely compared to other estimates derived from routine program data, in a capture-recapture estimation of the modeled population size (Abdul-Quader et al., 2014). The main challenge for community-based demographic surveillance is developing simple and sustainable data collection systems which can support weekly data collection and supervision, verification of data, and routine analysis (Ratnayake et al., 2020). Simple systems for field procedures and supervision from the Health and Demographic Surveillance Site (HDSS) network could be elaborated to support CHWs in estimating IDP populations (Cunningham et al., 2019). A better understanding of the perspectives and incentives for displaced individuals to be counted needs consideration. They may equate enumeration with identification by national authorities; this then requires assurance to the community that the act of population estimation by CHWs is done in aggregate and is not connected to registration of individuals and households. This type of social enquiry does not need to be reinvented; efforts are underway to merge traditional epidemiological methods for outbreak with social science enquiry in real-time (Carter et al., 2021). This aligned approach could also apply to population estimation to understand drivers and barriers to being counted.

The approaches discussed here require further conceptualization and validation using epidemiological, demographic, and social science methods. A promising area of research and learning is the integration of well-organized field estimation together with statistical modeling to fill the gaps in estimation in inaccessible areas or to predict movements using individual-based mathematical modeling (Checchi and Koum Besson, 2022; Suleimenova et al., 2017). More broadly, IDPs present as a highly vulnerable group and better collection of related data could compromise their protection, given the risk of misuse by actors to identify their locations and vulnerabilities, and the ever-present risk of data breaches. The leak of UNHCR's biometric registration data of Rohingya refugees is a natural extension of what could happen with such data (Human Rights, 2021). Alternatively, the very existence of IDP population data may be denied by some governments. Who has access to fine-level IDP data and how it can be harnessed operationally and for advocacy on behalf of the needs of the population remain key questions. International standards for preserving the confidentiality of IDP and refugee data exist and should be extended to any new and innovative data collection and analyses (Expert Group on Refugee and Internally Displaced Persons Statistics, 2018a, 2018b). This includes anonymizing the data through removal of names (de-identification), releasing only aggregated data, and enacting legal and administrative steps that regulate the confidentiality and sharing of data (Expert Group on Refugee and Internally Displaced Persons Statistics, 2018a, 2018b). Finally, a commitment to the real-time estimation of IDP population sizes and displacement flows would lead to a more accurate picture of need for humanitarian assistance. Operational agencies and academic partners can leverage these novel methods to find logistical arrangements which are effective for rapid estimation at the level of a crisis. As well, how expertise is resourced at the local and regional level should be debated across agencies (Spiegel, 2007). While UNHCR carries the responsibility for estimating forcible displacement under their mandate, IOM, and OCHA should continue to develop the human resource infrastructure to oversee estimation of IDPs at the level of the crisis, and nongovernmental and civil society organizations should be provided a clear role in community-facing data collection (Franklinos et al., 2021). Above all, humanitarian actors should advocate for resources to routinely estimate IDP population sizes as it remains the very basis for all humanitarian assistance.

Funding statement

RR is supported by a Doctoral Foreign Study Award from the Canadian Institutes of Health Research (award number DFS-164266). NA is supported by the Bill and Melinda Gates Foundation (grant number INV-035643). CD is supported by Innovative Methods and Metrics for Agriculture, Nutrition and Health Actions (IMMANA) Programme, funded by UK Foreign Commonwealth and Development Office (FCDO) (grant number 300654), and the Bill and Melinda Gates Foundation (grant number INV-002962). The funders did not have any role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abdelmagid N., Checchi F. Estimation of population denominators for the humanitarian health sector: guidance for humanitarian coordination mechanisms, 2018.

Abdul-Quader, A.S., Baughman, A.L., Hladik, W., 2014. Estimating the size of key populations: current status and future possibilities. Curr. Opin. HIV AIDS 9 (2), 107–114.

Acosta, R.J., Kishore, N., Irizarry, R.A., Buckee, C.O., 2020. Quantifying the dynamics of migration after Hurricane Maria in Puerto Rico. Proc. Natl. Acad. Sci. U. S. A. 117 (51), 32772–32778.

André, C., 2020. Why more data is needed to unveil the true scale of the displacement crisis in Burkina Faso and Cameroon. Expert Opinion. Internal Displacement Monitoring Centre. https://www.internal-displacement.org/expert-opinion/why-more-data-is-needed-to-unveil-the-true-scale-of-the-displacement-crises-in.

Baaees, M.S.O., Naiene, J.D., Al-Waleedi, A.A., et al., 2021. Community-based surveillance in internally displaced people's camps and urban settings during a complex emergency in Yemen in 2020. Confl. Health 15 (1), 54.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R., von Schreeb, J, 2011. 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 (8), e1001083.

Bharti, N., Lu, X., Bengtsson, L., Wetter, E., Tatem, A.J., 2015. Remotely measuring populations during a crisis by overlaying two data sources. Int. Health 7 (2), 90–98.

Boo, G., Darin, E., Leasure, D.R., et al., 2022. High-resolution population estimation using household survey data and building footprints. Nat. Commun. 13 (1), 1330.

Bowden, S., Braker, K., Checchi, F., Wong, S., 2012. Implementation and utilisation of community-based mortality surveillance: a case study from Chad. Confl. Health 6 (1), 11.

Carter, S.E., Ahuka-Mundeke, S., Pfaffmann Zambruni, J., et al., 2021. How to improve outbreak response: a case study of integrated outbreak analytics from Ebola in Eastern Democratic Republic of the Congo. BMJ Glob. Health 6 (8).

Checchi, F., Koum Besson, E.S, 2022. Reconstructing subdistrict-level population denominators in Yemen after six years of armed conflict and forced displacement. J. Migr. Health 5.

Checchi, F., Stewart, B.T., Palmer, J.J., Grundy, C., 2013. Validity and feasibility of a satellite imagery-based method for rapid estimation of displaced populations. Int. J. Health Geogr. 12, 4.

Checchi, F., Warsame, A., Treacy-Wong, V., Polonsky, J., van Ommeren, M., Prudhon, C., 2017. Public health information in crisis-affected populations: a review of methods and their use for advocacy and action. Lancet.

Cunningham, S.A., Shaikh, N.I., Nhacolo, A., et al., 2019. Health and demographic surveillance systems within the child health and mortality prevention surveillance network. Clin. Infect. Dis. 69 (Suppl 4), S274–S2S9.

Diggle, E., Welsch, W., Sullivan, R., et al., 2017. The role of public health information in assistance to populations living in opposition and contested areas of Syria, 2012-2014. Confl. Health 11, 33.

Ecopia. A.I., Maxar technologies. Digitize Africa Data — Building footprints. 2022. https://www.maxar.com/products/imagery-basemaps.

Expert Group on Refugee and Internally Displaced Persons Statistics. Technical report on statistics of internally displaced persons: current practices and reccomendations for improvement. Luxembourg, 2018.

Expert Group on Refugee and Internally Displaced Persons Statistics, International recommendations on refugee statistics. Luxembourg, 2018.

Franklinos, L.H.V., Parrish, R., Burns, R., et al., 2021. Key opportunities and challenges for the use of big data in migration research and policy. UCL Open Environ. 3.

Global Health Cluster, 2018. Programming in Access-Constrained Environments: Practical Guidance. WHO, Geneva, Switzerland.

Human Rights, Watch.. UN shared Rohingya data without informed consent: Bangladesh provided Myanmar information that refugee agency collected. https://www.hrw.org/news/2021/06/15/un-shared-rohingya-data-without-informed-consent.

Huynh, B.Q., Basu, S., 2020. Forecasting internally displaced population migration patterns in syria and yemen. Disaster Med. Public Health Prep. 14 (3), 302–307.

International Organization for Migration (IOM), 2021. Displacement Tracking Matrix: About. https://dtm.iom.int/about.

Koum Besson, E.S., Norris, A., Bin Ghouth, A.S., et al., 2021. Excess mortality during the COVID-19 pandemic: a geospatial and statistical analysis in Aden governorate, Yemen. BMJ Glob. Health 6 (3).

Leasure D.R., Kashyap R., Rampazzo F., et al. Ukraine Crisis: monitoring population displacement through social media activity 2022 (pre-print). SocArXiv Papers 2022.

Li T., Dejby J., Albert M., Bengtsson L., Lefebvre V. Detecting individual internal displacements following a sudden-onset disaster using time series analysis of call detail records arXiv [preprint] 2019.

Maas, P., Iyer, S., Gros, A., Franco, Z, González, JJ, Canós, JH, et al., 2019. Facebook disaster maps: aggregate insights for crisis response & recovery. In: Proceeding of the

16th Conference on Information Systems for Crisis Response and Management (ISCRAM). València.
Palotti, J., Adler, N., Morales-Guzman, A., et al., 2020. Monitoring of the Venezuelan exodus through Facebook's advertising platform. PLoS ONE 15 (2), e0229175.
Princeton School of Public and International Affairs. Assessing the use of call detail records (CDR) for monitoring mobility and displacement. https://www.migrationdataportal.org/resource/assessing-use-call-detail-records-cdr-monitoring-mobility-and-displacement.
Quinn, J.A., Nyhan, M.M., Navarro, C., Coluccia, D., Bromley, L., Luengo-Oroz, M., 2018. Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. Philos. Trans. A Math. Phys. Eng. Sci. 376 (2128).
Ratnayake, R., Tammaro, M., Tiffany, A., Kongelf, A., Polonsky, J.A., McClelland, A., 2020a. People-centred surveillance: a narrative review of community-based surveillance among crisis-affected populations. Lancet Planet. Health 4 (10), e483–ee95.
Ratnayake, R., Rawashdeh, F., AbuAlRub, R., et al., 2020b. Access to care and prevalence of hypertension and diabetes among Syrian refugees in Northern Jordan. JAMA Netw. Open 3 (10) e2021678-e.
Sphere, 2018. The Sphere Handbook: Humanitarian Charter and Minimum Standards in Humanitarian Response.
Spiegel, P.B., Checchi, F., Colombo, S., Paik, E., 2010. Health-care needs of people affected by conflict: future trends and changing frameworks. Lancet 375 (9711), 341–345.
Spiegel, P.B., 2007. Who should be undertaking population-based surveys in humanitarian emergencies? Emerg. Themes Epidemiol. 4, 12.
Suleimenova, D., Bell, D., Groen, D., 2017. A generalized simulation development approach for predicting refugee destinations. Sci. Rep. 7 (1), 13377.
Tatem, A.J., 2017. WorldPop, open data for spatial demography. Sci. Data 4, 170004.
Total population, both sexes combined (thousands). https://data.un.org/.
UNHCR, 2021. Global Trends: Forced Displacement in 2020. UNHCR, Geneva, Switzerland.