Ayumi Arai

Displacement and Disaster Indicators Derived from Mobile Positioning Data

Ayumi Arai*[1]; Véronique Lefebvre[2]; Kenth Engø-Monsen[3]; Sriganesh Lokanathan[4]

[1] The University of Tokyo, Tokyo, Japan, arai@csis.u-tokyo.ac.jp
[2] Flowminder Foundation, Stockholm, Sweden, veronique.lefebvre@flowminder.org
[3] Telenor Research, Fornebu, Norway, kenth.engo-monsen@telenor.com
[4] Pulse Lab Jakarta, Jakarta, Indonesia, sriganesh.lokanathan@un.or.id

Abstract:
There is increasing interest among policy makers in measuring hazardous events and disasters—data that can be useful in disaster contexts. National statistical systems collect basic data on population, economy, and other relevant topics necessary for disaster risk management. Official statistics play a crucial role in responding to the data demand to support preparedness and responses. Timely quality information on affected populations is important for formulating effective responses to disasters. Traditional data sources such as census and survey data are reliable but have limitations in measuring large-scale population movements due to logistical constraints in data collection. The ubiquitous use of cell phones provides opportunities to gain valuable insights into such movements. A call detail record (CDR), generated for all subscribers in real time, can provide longitudinal data for a large number of people.
In this paper, we introduce mobility indicators for populations affected by not only natural but also biological disasters such as epidemics. We demonstrate the key steps to generate statistical information for measuring displacement and impact on activity through country cases[1]. This includes institutional frameworks, data pipelines, and methodological approaches for producing statistical outputs. We also explain how the information is provided to responders and used for decision-making during a disaster.
However, these challenges and successes are still unique to an extent in that they were a result of accessing CDR data. Data access remains difficult in many instances as well as a key limitation in the use of CDR data for official statistics. We, therefore, suggest the use of alternative data sources, which are made available for supporting humanitarian activities by the private sector.

Keywords:
Mobile positioning data; call detail record (CDR); official statistics; disaster; displacement

[1] This paper summarizes the Handbook of Displacement and Disaster Statistics, developed by the members of the UN Committee of Experts on Big Data and Data Science for Official Statistics, Mobile Phone Data Task Team, Displacement and Disaster Subgroup.
The members of the subgroup: Flowminder, Indian School of Business, IOM, ITU, Positium, Pulse Lab Jakarta, Telenor Research, University of Tokyo* (lead), UNSD, World Bank.

Ayumi Arai

Displacement and Disaster Indicators Derived from Mobile Positioning Data

Ayumi Arai*[1]; Véronique Lefebvre[2]; Kenth Engø-Monsen[3]; Sriganesh Lokanathan[4]

[1] The University of Tokyo, Tokyo, Japan, arai@csis.u-tokyo.ac.jp
[2] Flowminder Foundation, Stockholm, Sweden, veronique.lefebvre@flowminder.org
[3] Telenor Research, Fornebu, Norway, kenth.engo-monsen@telenor.com
[4] Pulse Lab Jakarta, Jakarta, Indonesia, sriganesh.lokanathan@un.or.id

Abstract:
There is increasing interest among policy makers in measuring hazardous events and disasters—data that can be useful in disaster contexts. Official statistics play a crucial role in responding to the data demand to support preparedness and responses. Timely quality information on affected populations is important for formulating effective responses to disasters. Traditional data sources such as census and survey data are reliable but have limitations in measuring large-scale population movements due to logistical constraints in data collection. The ubiquitous use of cell phones provides opportunities to gain valuable insights into such movements. A call detail record (CDR), generated for all subscribers in real time, can provide longitudinal data for a large number of people.
In this paper, we introduce mobility indicators for populations affected by not only natural but also biological disasters such as epidemics. We demonstrate the key steps to generate statistical information for measuring displacement and impact on activity through country cases[1*]. This includes institutional frameworks, data pipelines, and methodological approaches for producing statistical outputs. We also explain how the information is provided to responders and used for decision-making during a disaster.
However, these challenges and successes are still unique to an extent in that they were a result of accessing CDR data. Data access remains difficult in many instances as well as a key limitation in the use of CDR data for official statistics. We, therefore, suggest the use of alternative data sources, which are made available for supporting humanitarian activities by the private sector.

Keywords:
Mobile positioning data; call detail record (CDR); official statistics; disaster; displacement

1. Introduction:

Role of official statistics in response to increasing data demand

There are increasing policy interests in measuring hazardous events and disasters—data that can be useful for disaster risk management. This has been addressed in three high-level policy frameworks: the Sendai Framework for Disaster Risk Reduction 2015–2030 (UNDRR 2015), 2030 Agenda on Sustainable Development (United Nations 2015b), and Paris Agreement (United Nations 2015a); these are closely related to the Sustainable Development Goals, particularly 3 and 11. Official statistics play a crucial role in responding to this data demand (United Nations 2019).

[1*] This paper summarizes the contents of the Handbook of Displacement and Disaster Statistics, developed by the members of the UN Committee of Experts on Big Data and Data Science for Official Statistics, Mobile Phone Data Task Team, Displacement and Disaster Subgroup. The members of the subgroup: Flowminder, Indian School of Business, IOM, ITU, Positium, Pulse Lab Jakarta, Telenor Research, University of Tokyo* (lead), UNSD, World Bank.

Existing displacement and disaster indicators and data gaps
The extent to which populations are affected by disasters, including those who are evacuated and relocated from their homes to nearby city slums and rural areas, has not been well defined. Traditional data such as census and survey data are built on rigorous methodologies and widely used for a long time; these are reliable resources for understanding affected populations (Fussell *et al.* 2014) but have limitations in measuring large-scale population movements due to logistical constraints in data collection (Brown *et al.* 2001). It has been increasingly acknowledged that there are data gaps in examining the impact of disasters (Ager *et al.* 2014).
Relevance of using cell phone data in disaster contexts
Timely quality information on affected populations is crucial for formulating effective responses to disasters (UN Global Pulse 2014). The ubiquitous use of cell phones provides opportunities to gain valuable insights into large-scale population movements (Wesolowski *et al.* 2012) (Wilson *et al.* 2016). A call detail record (CDR) is information on a mobile network event, such as a call or short message service (SMS)—it is generated at every network event and includes its time and location information. Thus, with data being generated for all subscribers in real time, longitudinal data can be provided for a large number of populations. Previous studies have shown the usefulness of CDR data for tracking population movements following a disaster. For example, in Haiti, CDR data were used to measure population mobility and displacement patterns in the aftermath of the 2010 earthquake. The data covered approximately 63% of the total subscribers and 90% of the inhabited areas, and the result showed the high correspondence of geospatial distribution of population movements between CDR data and large-scale retrospective survey data (Bengtsson *et al.* 2011).

2. Data characteristics to be considered
Impact of biases on statistical inference
How the data represent the general population's behaviors largely depends on the type of problems to be solved and the nature of the statistical inference drawn from the data (Tam and Clarke 2015). In disaster contexts, selection and measurement biases must be taken into account for the following reasons. Selection bias is caused because the CDR data include only cell phone users. Disasters adversely affect the poor and more vulnerable people who are less likely to have cell phones. Moreover, policymakers and responders are also more interested in the poor, elderly, children, and women, who are unlikely to have cell phones. If mobile penetration rate is high enough, CDR data could potentially cover such populations to an extent unless their mobility patterns are substantially different from the cell phone users' during a disaster (Bengtsson *et al.* 2011). Phone surveys, while limited in assessing representativeness of the general population in terms of mobility, can help provide phone users' demographics with respect to their mobility and frequency of phone use, as well as validate analytical methods for detecting migration and displacement. Measurement bias is caused because a CDR is generated only when a network event occurs. The extent to which the CDR data represent the actual mobility patterns of cell phone users largely depends on the frequency of records. For instance, detailed movements of users are unobservable if they do not use their phone frequently (Deville et al. 2014). Measurement bias's impact could be low when studying long-term location and mobility (Wilson et al. 2016) but extensive when studying hourly or daily patterns. This is because individual trajectories are characterized by statistical regularity, which can be explained by preferable returns to a few locations and occasional explorations to other locations (González, *et al.* 2008).
Data privacy and protection
The displacement and disaster indicator introduced in this paper is CDR aggregates. Data aggregation is part of the pre-process, which enables the data producer to maintain privacy (Buckee and Engø-Monsen 2016). It is important to aggregate the data at certain geographic levels and ensure that the aggregates include enough population in groups (Sweeney 2000); moreover, combining multiple levels of aggregation in space and time can also provide further information. In addition to aggregation, dropping small counts also enables privacy

protection. In general, grid cells including up to 20 people should not be shared (Sweeney 2000), while GSM Association (GSMA) guidelines recommend dropping all counts below 15 subscribers.

3. Institutional and analytical setups for generating indicators

Institutional framework supporting a timely response

It usually takes time to reach a consensus on the use of CDR data, including the modality of data sharing. Having an institutional framework allows the government to quickly extend development objectives for partnerships for responding to emergency situations. Even if an existing framework or data pipeline is not perfect for producing statistical products, it can lower the response burden of both the mobile network operator (MNO) and regulator while protecting their information.

Table 1 presents country cases in which existing partnerships enabled timely responses during disasters. In Nepal, the first report on population inflows to each region was released 13 days after the earthquake—this was made possible by the existing collaboration framework, which aimed to support preparedness and response to humanitarian disasters through analysis of CDR data (Wilson *et al.* 2016). In Vanuatu, at the time of volcanic eruptions in 2018, there was already an institutional framework in place for analyzing the effects of previous disasters—it enabled timely insights into population displacement for the government as well as humanitarian agencies (UN Global Pulse 2014). In The Gambia, the existing framework explored the use of CDR data to create an evidence base for policy and project design in the context of economic and social development; the spread of COVID-19 has prompted interest in internal mobility, altering the development objective of this partnership (Arai *et al.* 2021).

Table 1. Country cases where institutional frameworks supported timely responses

	Existing framework	Benefits and reduced burden
Nepal	Collaborative framework established six months before the earthquake (MNO, Flowminder)	● Supported the response to the 2015 Gorkha earthquake ● Infrastructure set up quickly and first report prepared within two weeks after the earthquake
Vanuatu	Analysis of the 2017 tropical cyclone and volcano ongoing (MNO, UN Global Pulse)	● Supported the response to the 2018 volcanic eruptions ● Data access and code in place, enabling to provide timely insights
The Gambia	Analysis of migrations ongoing since 2019 (NSO, regulator, MNOs, World Bank, University of Tokyo)	● Supported the response to COVID-19 in 2020 ● Data pipeline already in place and ready for generating indicators for monitoring and planning

Data pipelines strengthening preparedness

After the onset of a disaster, when the situation is volatile and quick decision-making is required, growing data demands can easily overwhelm the analytical and processing capacity. In such a case, it is important to have an analytical pipeline that can provide information products such as aggregate statistics and standardized indicators; it could lower the response burden and advance the analytical objectives (Kishore *et al.* 2020).

Table 2 presents country cases where data pipelines provided actionable information for decision-making, which contributed to timely analysis based on CDR data. Sierra Leone, Guinea, and Liberia collaborated to develop a data pipeline, which facilitated timely information exchange on cross-border movements among the neighboring countries (ITU n.d.). In Turkey, a real-time analytics platform was developed to provide analytical insights to the government disaster response team; it helped zero in on the number of people affected by disasters nationwide (GSMA 2019). In Mozambique, a semi-automated software tool was installed on the premise of the ICT regulator to detect disaster-driven displacements and returns, providing results in a practical and timely manner; in this case, past mobile data was analyzed to assess the sensitivity and accuracy of the displacement monitoring system for future use. This toolkit was also installed in Ghana to support the COVID-19 response effort.

Table 2. Country cases where data pipelines strengthened preparedness

	Partnership	Benefits of data pipelines
Sierra Leone, Guinea, Liberia	Regulators, MNOs, ITU, University of Tokyo	● Facilitated timely information exchange on human mobility patterns, including cross-border movements during the Ebola epidemic
Turkey	MNO, GSMA	● Developed a system (Galata) to deliver real-time analytics during earthquakes ● Ensured functionality through testing and field exercises
Mozambique	Regulator, disaster management agency, Flowminder	● Generated aggregated data for identifying displacement ● Used an open-source toolkit (FlowKit) for data management
Norway	MNO, NIPH, University of Oslo, Norwegian Computing Centre	● Operated daily for data extraction from the MNO, collection of health data, and analytics ● Supplied the Norwegian Institute of Public Health (NIPH) with data daily to support COVID-19 response efforts

4. Displacement and disaster indicators

Disaster displacement can have a devastating impact on individuals and communities by disrupting their lives. To identify displaced populations and evacuees, it is critical to have information on the exact number of people in a specific area at a particular time, as well as on mobility across regions. This enables the estimation of the number of people affected and what the subsequent changes are. Table 3 presents country cases where CDR data were used for detecting displacement and examining the impact on activity in disaster contexts; it summarizes the processes of computing indicators and examining their statistical relevance.

Detecting displacement

In Bangladesh, changes in the number of active SIMs were used as proxy for population transitions after Cyclone Mahasen in 2013; this highlights the benefit of CDR data, which enables us to measure small movements across vast areas—these are extremely difficult to measure using traditional survey-based approaches (Lu *et al.* 2016) because such populations spread easily across large areas (Fussell *et al.* 2014). In Nepal, home locations were used to build transition matrices (origin-destination matrix), describing the countrywide mobility between two points in time. Individual home locations were determined by calculating the modal daily location over a certain period at the district level—a user was counted as displaced if they had spent at least seven consecutive days away from their pre-earthquake home location in a two-week period after the 2015 earthquake. This case demonstrates how cellular network data can be used to detect displacements, including returning residents, based on changes in home locations (Wilson *et al.* 2016).

For these indicators, it should be noted that the number of people proxied by active SIMs is a function of phone use and the number of active subscribers. The estimated results can reflect changes in cell phone use instead of those in population numbers. In addition, the value presented in a transition matrix is the number of people moving between two regions, which highly depends on the spatial and time scales used for computation (Kishore *et al.* 2020) as well as the definition of locations used as the origin and destination. For capturing long-distance travel, larger time windows are useful, as it typically takes longer to move from an origin region to the destination one. When only overnight locations are used and other trips such as commuting are excluded from the matrix, it enables the identification of long-term relocation. On the other hand, to understand short-term mobility fluctuations, such as those within a day, smaller time windows are useful.

Examining the impact on activity

Changes in the magnitude of mobility within a population can be a useful indicator for examining the impact of disasters and interventions—it examines the vector of travel, generated from every consecutive pair of cell towers of the CDR; the value is a function of cell tower density, and thus, the computed result highly depends on the area size covered by a cell tower. For instance, in rural areas where cell tower density is sparse, short-distance travel cannot be accounted for; travel beyond the boundary of distant cell towers can indicate a long-distance trip, which might not represent reality. In addition, it can be heavily affected by cell tower transitions, which generate “travels” that do not actually occur. Clustering neighboring cell towers can mitigate the transitions’ impact.

Table 3. Processes of computing indicators and examining their relevance

Country	Partnership	Steps taken for producing indicators and methods examining statistical relevance
Bangladesh (cyclone, 2013)	MNO, Flowminder	● Changes in the number of active SIMs as proxy for population transitions during and after weather events ● The spatial distribution of active SIMs compared with that as per the 2011 census (r=0.948 (p<0.001))
Nepal (earthquake, 2015)	MNOs, Flowminder	● Changes in home locations, estimated as the modal location of every individual, used for analyzing origin, destination, and relocation size ● Scaled estimates of total inflows per district compared with the census population data for the baseline period, with the result showing a close match
Haiti (earthquake, 2010)	MNOs, Flowminder	● Net outflows, computed from the number of active SIMs, used for estimating the distribution of population displaced from the capital ● Estimates of geographical distributions across Haiti similar to those derived from the retrospective study
Indonesia (earthquake and tsunami)	MNOs, Pulse Lab Jakarta, IOM	● The number of subscribers (weekly aggregates) used to identify displacement
The Gambia (COVID-19)	NSO, regulator, MNOs, World Bank, University of Tokyo	● Distance between the cell towers of consecutive data points used as proxy for distance traveled per person—aggregated by home region ● Population density by home region for the baseline period compared with that as per the census, with the result showing high correspondence
Sierra Leone (COVID-19)	Sierra Leone Directorate of Science, Technology and Innovation, MIT, Flowminder	● Distance between the locations visited by the subscribers within 24 hours (median across subscribers) used as proxy for distance traveled ● Total number of network events (records) used for checking for potential bias in the data following a pricing change by the operator

5. Conclusions

In this paper, we presented the key steps to generate displacement and disaster indicators, including methodological approaches for producing statistical outputs, institutional frameworks, and data pipelines. However, the related challenges and successes are still unique to an extent in that they were a result of accessing CDR data. In many instances, securing data access remains difficult, requiring a lot of effort, time, and coordination; it remains a key limitation in the use of CDR data for official statistics. In disaster contexts, alternative data sources are made available to support humanitarian activities by the private sector; these provide access to anonymous privacy-compliant location-based data for humanitarian initiatives related to human mobility.

References:

Ager, A. et al. 2014. “Strengthening the Evidence Base for Health Programming in Humanitarian Crises.” *Science* 345(6202): 1290–92.

Arai, Ayumi, Erwin Knippenberg, Moritz Meyer, and Apichon Witayangkurn. *forthcoming*. The Hidden Potential of Call Detail Records in The Gambia. *Data & Policy*.

Bengtsson, Linus et al. 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): 1–9.
Brown, V. et al. 2001. “Rapid Assessment of Population Size by Area Sampling in Disaster Situations.” *Disasters* 25(2): 164–71.
Buckee, Caroline O, and Kenth Engø-Monsen. 2016. *Mobile Phone Data for Public Health: Towards Data-Sharing Solutions That Protect Individual Privacy and National Security*.
Deville, Pierre et al. 2014. “Dynamic Population Mapping Using Mobile Phone Data.” *Proceedings of the National Academy of Sciences*.
Fussell, Elizabeth, Katherine J. Curtis, and Jack DeWaard. 2014. “Recovery Migration to the City of New Orleans after Hurricane Katrina: A Migration Systems Approach.” *Population and Environment* 35(3): 305–22.
González, Marta C., César A. Hidalgo, and Albert László Barabási. 2008. “Understanding Individual Human Mobility Patterns.” *Nature* 453(7196): 779–82.
GSMA. 2019. *Utilising Real-Time Mobile Analytics to Inform Emergency Disaster Response in Turkey: Turkey and Natural Disasters*.
ITU. “Big Data for Development: Preventing the Spread of Epidemics.” Retrieved May 29, 2021.
Kishore, Nishant et al. 2020. “Measuring Mobility to Monitor Travel and Physical Distancing Interventions: A Common Framework for Mobile Phone Data Analysis.” *The Lancet Digital Health* 7500(20).
Lu, Xin et al. 2016. “Unveiling Hidden Migration and Mobility Patterns in Climate Stressed Regions: A Longitudinal Study of Six Million Anonymous Mobile Phone Users in Bangladesh.” *Global Environmental Change* 38(May): 1–7.
Sweeney, L. 2000. *Uniqueness of Simple Demographics in the U.S. Population*.
Tam, Siu Ming, and Frederic Clarke. 2015. “Big Data, Official Statistics and Some Initiatives by the Australian Bureau of Statistics.” *International Statistical Review* 83(3): 436–48.
UN Global Pulse. 2014. *Using Mobile Phone Activity for Disaster Management during Floods*. https://www.unglobalpulse.org/document/using-mobile-phone-activity-for-disaster-management-during-floods/.
UNDRR. 2015. *Sendai Framework for Disaster Risk Reduction 2015-2030*. Retrieved May 21, 2021.
United Nations. 2015a. *Paris Agreement*.
———. 2015b. *Transforming Our World: The 2030 Agenda for Sustainable Development*.
———. 2019. *Recommendations on the Role of Official Statistics in Measuring Hazardous Events and Disasters*.
Wesolowski, Amy et al. 2012. “Heterogeneous Mobile Phone Ownership and Usage Patterns in Kenya.” *PloS one*.
Wilson, Robin et al. 2016. “Rapid and near Real-Time Assessments of Population Displacement Using Mobile Phone Data Following Disasters: The 2015 Nepal Earthquake.” *PLoS Currents* 8(Disasters).