4.2 Limitations and caveats

Representativeness in disaster context

How the data represent the general population's behaviors largely depend on the type of problems to be solved and the nature of the statistical inference drawn from the data (Tam and Clarke 2015). The characteristics of CDR data, including benefits and challenges for using the data for official statistics, are described in detail in the methodological guide of dynamic population.

Selection bias may arise because CDR data include only mobile phone subscribers. There are differences in phone and/or subscriber-identification-module (SIM) holding across geography and socioeconomic groups. The elderly and small children are less likely to be representative while it is biased towards males, higher income groups, and certain age groups (Wesolowski, Eagle, Noor, Snow, & Buckee, 2013) (Deville et al., 2014). Disasters adversely affect the poor and more vulnerable people who are less likely to have cell phones in some areas. Moreover, policymakers and responders are also more interested in the poor, elderly, children, and women, who are unlikely to have cell phones.

In the case of the 2010 earthquake of Haiti, 20% of the pre-earthquake population was estimated to have left the capital city. Geographic distribution of population outflows from the city corresponded well with results from a large retrospective survey, which included a representative sample of 2,500 households in the city. Among the survey sample, approximately 3,000 persons had left the city following the earthquake. The similarities in patterns from two datasets could be due to a widespread use of mobile phones; the movement of nonmobile phone users together with mobile phone users, and also the similarities in movement patterns of users and nonusers of mobile phones. The result indicates that CDR data can reflect the mobility patterns of people who are less likely to have phones to an extent if the mobile penetration rate is high enough (Bengtsson et al. 2011). As of 2010, the mobile penetration rate of Haiti was more than 40% (ITU, n.d.-b) and that of the capital city is expected to be much higher. The CDR data were provided by a mobile network operator (MNO) who had the largest market share at that time.

Measurement bias may also arise 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 (Couper, 2013). For instance, detailed movements of users are unobservable if they do not use their phone frequently (Deville et al. 2014). Similar bias could occur when analysis is conducted for a part of subscribers or CDR data after filtering those who have very limited number of records; results could be affected by the filtering process due to potential correlations between individual mobility and device usage (Liu, Sui, Kang, & Gao, 2014). Measurement bias’s impact could be low when studying long-term location and mobility, e.g., when home locations are calculated using the modal value of multiple consecutive daily locations (Wilson et al. 2016) but extensive when studying hourly or daily patterns (with only those frequent users being represented) and when studying changes after shocks which may affect both mobility and phone usage of all subscribers (Flowminder report on Covid in Sierra Leone). 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 (Song et al. 2010) (González et al. 2008). However, metrics such as 'distance travelled' are less sensitive to frequency of phone usage than metrics such as subscriber presence and daily flows.

In addition, multiple-SIM holding is quite common in low- and middle-income countries, which affects the representativeness of the data. Also, phone-sharing practice could create noise (tangled signals) in estimated mobility patterns (Wesolowski, Eagle, Noor, Snow, & Buckee, 2012), and so could changes in ownership of phones and SIM cards over time. These biases need to be considered when the result of CDR data analysis is interpreted. The biases can be mitigated using other data sources (Wesolowski et al., 2013).

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 (Wilson et al. 2016) as well as validate analytical methods for detecting migration and displacement; information on market share is important for examining representativeness, including the socio-economic characteristics of mobile phone users, e.g. the wealthier, poorer, younger, compared with the general population. Information on geospatial coverages could be useful for examining the geographic representativeness. Understanding how adequately or disproportionately the data represent rural communities is crucial if the population under study are in rural areas because CDR data would be most representative of urban populations (Kishore et al., 2020).