

Bias estimates for MPD-derived mobility statistics in low-and middle-income countries

ROLAND HOSNER*, ZACHARY STRAIN-FAJTH, VÉRONIQUE LEFEBVRE

Flowminder Foundation, Norrskén House, Birger Jarlsgatan 57C, 113 56 Stockholm, Sweden

*roland.hosner@flowminder.org

Received 5 July 2025; revised 8 Oct 2025; accepted XX Month XXXX; posted XX Month XXXX (Doc. ID XXXXX); published XX Month XXXX

ABSTRACT

Based on original and recent household and telephone surveys in the Democratic Republic of the Congo (DRC), Haiti and Ghana, we estimate biases of mobility and migration statistics derived from mobile phone data (MPD). We show that indicators of short- and long-term internal mobility differ greatly between mobile phone users and non-users, where phone users show higher mobility on average. These differences translate into biases when statistics are derived from mobile phone data and population inferences are made without adjustments. These findings are not only relevant for mobility research but also have wider implications for the use of Mobile Phone Data and other Big Data for analyses of mobility and migration.

INTRODUCTION

The interest in and use of Big Data for the analysis of human mobility has grown rapidly in recent years. Particularly in low- and middle-income countries (LMICs), the promise of utilising highly granular data for official statistics, public service provision and for humanitarian and development purposes is driving analyses of relatively new data sources such as MPD for statistics. Use cases range from routine mobility reports (1) to disaster response (2), epidemiological studies (3) and the study of rare events such as mobility in the wake of earthquakes and other natural hazards (4). There is also growing awareness of and consensus among researchers that biases inherent in such novel data, particularly stemming from the non-probability nature of these samples, need to be analysed, and where feasible, adjusted for (5-7). However, the sample selection mechanism cannot be identified from MPD alone, socio-demographic information is often incomplete or missing, so that standard data sources – survey, census and register data – have to be used in combination with MPD to assess and mitigate biases.

The work we present here covers the assessment and is aimed at highlighting the differences in mobility and migration indicators between mobile phone users and non-users that translate into biases when only MPD are used for target statistics. In LMICs, direct measures of the target statistics are often lacking, and MPD-derived statistics are often more granular in time and space than available statistics at higher temporal and spatial levels, making it

impossible to calculate exact bias statistics. Therefore, we present bias estimates based on survey data.

DATA

For the DRC, the data of the 2021 microcensus were collected through face-to-face household interviews in seven provinces. For Ghana, we were able to submit questions to an ongoing household survey, the Annual Income and Household Expenditure Survey (AHIES) 2022, covering all regions of Ghana. For Haiti, the household survey data used for analyses stem from the DHS data collection 2016/2017.

RESULTS

For all three countries we find significant differences in mobility between mobile phone users and non-users, across a range of mobility indicators. In Ghana, the differences are generally larger than in Haiti and the DRC. Short-term mobility, such as spending one or several months during the year away from home, is more common among mobile phone users than non-users in both Ghana and Haiti. A more permanent and rarer type of mobility - having changed the place of residence in the previous six months - showed smaller differences between mobile phone users and non-users in both the DRC and Ghana. Long-term mobility, measured as any change of place of residence over the life course, is more prevalent overall, and significantly more common among mobile phone users in all three countries (see table 1). A smaller difference can be seen for the DRC, while in Haiti, the difference is much larger, with half of mobile phone users ever having internally migrated, but only about four out of ten non-users. The largest difference is seen for Ghana, with half of mobile phone users ever having migrated internally, but only a third of non-users.

Table 1. Mobility differences between mobile phone users and non-users in DRC, Haiti and Ghana

Mobile phone use	Lifetime internal migrant		
	DRC	HTI	GHA
Non-users	40.1%	38.3%	34.7%
Mobile phone users	42.9%	51.3%	50.9%
p-value ^a	0.031	0.000	0.000
Observations	9,833	25,033	30,774

^a Wald tests for differences of means

Sources: DRC Microcensus 2021, Haiti DHS 2017, Ghana AHIES 2022 Q2

Across the three countries, we analysed a total 15 indicators of mobility and migration and found significant effects ($p < 0.1$) of mobile phone use for 13 of these indicators in bivariate regression models in overall models, of which 10 were highly significant ($p < 0.01$). All 13 indicators show higher mobility of mobile phone users in these overall models, where all respondents per country are included. The bivariate effects were partly explained when controlling for socio-demographic variables and region (administrative level 1), but 12 indicators remained significant and 11 highly significant also in the multivariate analyses.

Moving from mean tests to bias estimation, whenever general population parameters are estimated directly from samples of mobile phone users (excluding non-users in general or also users of MNOs not covered), the bias of such estimates is the difference between these estimates and the true population parameters. The population mean and therefore the bias are usually not known, but we can estimate this bias from survey estimates, by examining the difference between the mean estimate among phone users in the survey data and the grand mean estimate in the survey data, covering both phone users and non-users:

$$\text{bias}(\hat{\mu}_{SD-MP}) = \hat{\mu}_{SD-MP} - \hat{\mu}_{SD}$$

If and when the parameter of interest differs between mobile phone users and the general population, there is bias in the direct MPD-derived estimate.

Now taking Ghana as an example and using bias in % points as a measure for dichotomous indicators, biases across five mobility indicators and 16 regions (admin1) ranged between -1% and +18%. In other words, the mobility of mobile phone users was often higher than for the general population, resulting in a positive bias. The spread of biases on the district (admin2) level is much larger than that of biases on the regional level. Across all indicators and 177 districts, biases range from -6 % points to +31 % points (fig. 1). This shows that on lower sub-national levels, biases vary more substantially, and smaller higher-level biases may mask variation on lower levels.

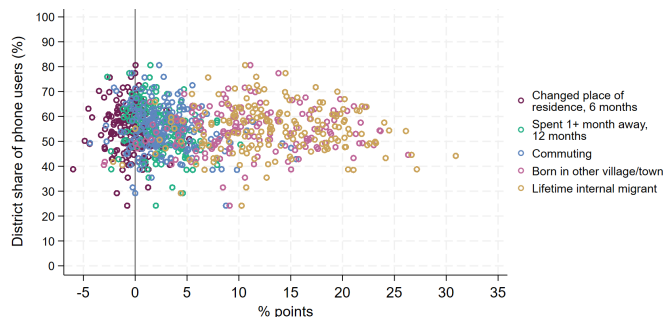


Fig. 1. Bias estimates for five mobility indicators across 177 Ghana districts. Data Source: Ghana Annual Income and Household Survey 2022, quarter 2

Biases for more permanent forms of mobility show substantially larger variation in our data. However, this might be due to the so-called floor effect: when proportions are small, variation of these proportions between groups is statistically curtailed.

DISCUSSION

In summary, mobile phone use has a significant impact on mobility and migration statistics derived from MPD, which translates into biased estimates. Such biases may be small on the national level but become larger on sub-national levels. While acknowledging that it may not be possible to eliminate but only mitigate bias, future research should address these statistical quality issues by triangulating MPD and standard data sources such as surveys and providing estimates that are bias-adjusted, so that MPD-derived statistics can be used for analyses and practical applications, for example by leveraging integrated estimates (8) or modelled estimates (9).

In practical terms, when a statistic is biased by 10% points or more, this may have serious consequences for the interpretation of results and the applied use case. Health workers may allocate scarce resources in places that are more or less populated than indicated by the MPD-derived statistics, for example. As such, accuracy of statistics for the service provision, for humanitarian and development sectors is equally important as for official statistics.

References

1. Ghana Statistical Service. Population Mobility and Distributions in Ghana in 2024. 2025.
2. Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Med.* 2011; 8(8).
3. Kostandova N, Prosperi C, Mutembo S, Nakazwe C, Namukoko H, Nachinga B, et al. Comparing and integrating human mobility data sources for measles transmission modeling in Zambia. *PLOS Global Public Health.* 2025; 5.
4. 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. 2019.
5. Milusheva S, Björkegren D, Viotti L. Assessing Bias in Smartphone Mobility Estimates in Low Income Countries. In: *Proceedings of 4th ACM SIGCAS Conference on Computing and Sustainable Societies.* 2021. p. 364–78.
6. Pestre G, Letouzé E, Zagheni E. The ABCDE of Big Data: Assessing Biases in Call-Detail Records for Development Estimates. *World Bank Economic Review.* 2020;34: 89–97.
7. Gallotti R, Maniscalco D, Barthelemy M, De Domenico M. Distorted insights from human mobility data. *Commun Phys.* 2024; 7(1).
8. Kim JK, Tam SM. Data Integration by Combining Big Data and Survey Sample Data for Finite Population Inference. *International Statistical Review.* 2021; 89(2): 382–401.
9. Zhang LC, Necula M, Filippini D, Simeoni G. Methodologies and open source tools for integrating MNO and non-MNO data sources. Preliminary report on methodologies. 2024.

Funding. The work was funded by the Hewlett Foundation

Data availability. The survey microdata underlying the results presented in this paper are available from Ghana Statistical Service upon request, from the DHS website and from the DRC Institut National de la Statistique upon request.

Disclosures. The authors declare no conflicts of interest