FLOWMINDER.ORG

Methodology report | Phone survey

Ghana Mobile Phone Survey 2022

Ghana Mobile Phone Survey 2022

Methodology Report

1. Overall aim and purpose

1.1. Project context

The Ghana phone survey 2022 was implemented as part of **phase 2 of the Data for Good Partnership** in Ghana - a public-private partnership and not-for-profit initiative to support evidence-based decision-making for the wellbeing of all in Ghana, between Vodafone Ghana (now Telecel Ghana), the Flowminder Foundation, and Ghana Statistical Service (GSS). The project received funding support from Vodafone Foundation (now Telecel Foundation) and the William and Flora Hewlett Foundation. The aim of the project was to use pseudonymised and aggregated mobile operator data to estimate population distributions and trends of population movements across Ghana for development and humanitarian purposes.

The **objective** of the Ghana phone survey 2022 was also to validate Flowminder's methods of production of mobility estimates based on Call Detail Records (CDRs), particularly of key concepts (home location, mobility, migration), and to learn about the socio-demographic structure of phone users in Ghana.

1.2. Project partners

The analysis team at **Flowminder** created the research design of the phone survey and oversaw all stages of its implementation, from drafting the research questions, the sampling design, the questionnaire, selecting and commissioning the implementing organisation, monitoring of data quality, data processing and cleaning as well as data analysis and reporting.

Ghana Statistical Service (GSS) was selected as the implementing organisation and was responsible for questionnaire programming and testing, selection and training of interviewers as well as data collection, quality monitoring and initial data cleaning. GSS also contributed to the adaptation of the final questionnaire to the Ghana context.

The gross sample for the telephone survey was created by **SampleSolutions**, based in the Netherlands.

The **Ethics Committee for the Humanities of the University of Ghana** reviewed and approved the survey design and questionnaires.

1.3. Content and topics

The objective of Flowminder's Ghana phone survey 2022 was to understand more about the mobility of the Ghana population as well as differences in demographics and mobility of various telephone users' groups, in order to inform Flowminder's estimates of internal mobility in Ghana.

In particular, the survey data are to be used for

- the assessment of measurement biases in CDR data arising from SIM sharing and multiple SIM use
- the assessment of household mobility
- the assessment of the demographic structure of phone users and various groups of users
- the validation of home locations as conceptualised from CDR data and
- the validation of indicators for mobility and migration.

Therefore the questionnaires covered the following topics

- Module I - Pre-call information
- Module S - Screening questions
- Module X - Home location and gender
- Module R - Mobile phone use
- Module M - Migration and mobility
- Module A - Individual socio-economic status
- Module H - Household socio-economic status
- Module E - End
- Module Z - Interviewer questionnaire

2. Legal basis and data protection

The initial lawful basis for data processing was legitimate interests pursued by Flowminder (Article 6(1)(f), GDPR) (Article 20(1)(e), Data Protection Act of 2012, Ghana). These legitimate interests are the following: improving access to health and social services, health service delivery planning, disaster management planning and response, marking progress towards the Sustainable Development Goals, as well as scientific research.

This initial lawful basis was required to enable a first contact to respondents. Once that contact was established, at the start of each call the interviewers introduced themselves, GSS and Flowminder (as the commissioning organisation), outlined the purpose and goals of the survey, the project context, highlighted the confidential nature of all data collected and that participation was voluntary. It was also outlined that the respondent could skip questions and that they could stop the interview at any time. Furthermore, respondents were informed that the data would be shared with Flowminder Foundation and GSS. Informed consent was secured for all interviews used for analysis.

All respondents who completed the interview were also referred to the applicable privacy notice on Flowminder's website, outlining respondents' rights in more detail: www.flowminder.org/survey.

3. Methodology

3.1. Mode of data collection

Data was collected through individual telephone interviews with Ghana phone users, i.e., **Computer-Assisted Telephone Interviews (CATI)**.

3.2. Statistical concepts, definitions and classifications

The development of the English source questionnaire and the indicators used was informed by the overall aim of the project - analysing mobility for statistical and planning purposes - and Flowminder's related analyses needs.

The indicators for **individual and household socio-economic status** for modules A and H were taken from the Demographic and Health Surveys (DHS) Household Questionnaire for phase 8.

Several indicators on **phone use** (module R) and **mobility** (module M) were taken and adapted from the 2021 DRC phone survey, which was also designed and coordinated by Flowminder.

For the classification of attempted, partially completed and completed codes, the disposition codes suggested by AAPOR were used.

3.3. Statistical population

Target population: was defined as all mobile phone users in Ghana who were using a phone number of a Ghana Mobile Network Operator (MNO) and who (a) were the main user of that SIM card, (b) were resident in Ghana at the time of the call, and (c) were at least 15 years of age.

The existing network prefixes at the time of data collection were:

- Airtel: +233 26, +233 56
- Glo: +233 23
- MTN: +233 24 , +233 54, +233 55, +233 59
- Tigo: +233 27, +233 57
- Vodafone: +233 20, +233 27

Primary Sampling Units were individual phone users who used a SIM with a phone number from a Ghana Mobile Network Operator.

Geographic coverage: Ghana, national level. The data include respondents from all 16 regions of Ghana.

Reference period: Interviews were conducted **between 25 November and 11 December 2022**. Data collection took 18 days.

3.4. Sampling and sample size

The survey was designed as a **cross-sectional stratified simple random sample telephone survey**.

The design foresaw **stratification by district**, with target net sample size of n=16 to n=23 interviews per district. However, this design was not fully implemented, with sample size per district ultimately ranging from n=1 to n=112. For four districts (Pru West, Krachi West Municipal, Sawla-Tuna-Kalba, Ketu North Municipal) no interviews could be completed. The median subsample size per district was n=26.

Random digit generation was used to create a sample of Ghana mobile phone numbers. For this purpose, SampleSolutions was subcontracted by Flowminder as a technical partner.

SampleSolutions provided a **gross sample of 11,000 Ghana mobile phone numbers**. The process to retrieve that gross sample involved multiple steps. In a first step, based on a potential total of existing mobile phone numbers related to prefix codes of providers (MNOs), a potential sampling frame was calculated. This potential sampling frame was much larger than the number of actual mobile phone numbers in use and only served the purpose of creating an initial sample. This pre-screening sample was screened via a Home Location Register (HLR) lookup - a check for the authenticity and activity status of a cell phone number - and via checks against SampleSolution's register of business numbers (B2B). Out of these numbers, a random subset of 11,000 numbers was selected and forwarded to GSS for interviewing. Table 1 shows the total and MNO-specific counts for the gross sample.

From this overall gross sample of 11,000 Ghana cell phone numbers, **8,452 numbers were called**, and a total of 7,913 called working numbers were recorded. The net sample (complete and partial interviews that can be used) consists of **4,974 interviews** (see table 1).

The two main reasons for unsuccessful calls were unanswered phone calls (1,752) and refusals (717). Unanswered phone calls comprised situations where there was a ringtone but the call was not answered, situations where an automated message informed the caller that the phone was currently switched off, and situations where the call did not pass (e.g. no ringtone). The **eligibility** of these numbers remains **unclear** (disposition code 3.0, U). Only in cases where there was an explicit message that the number dialled was incorrect, not assigned or not in service it was marked as a non-working number. **Refusals** (code 2.11, R) comprise all situations where the respondent explicitly refuses or declines to be interviewed. If the interview was aborted prematurely or the respondent gave valid answers to less than 50% of all applicable questions, that interview was marked as a **break-off** (code 2.12, R). Such situations were very rare (7 cases).

A comparatively small number of cases (713) were **not eligible** (code 4). Mainly these were non-working or disconnected numbers (539), cases where the district quota was filled (130), and a smaller number of out-of-sample cases (44), namely persons not living in Ghana, minors under 15 and business numbers.

Non-contacts (code 2.2, NC) include eligible cases where there was a conversation on the phone, but the interview was not started for reasons other than refusals; these cases were also relatively rare (248), for example when call-backs were agreed but the respondent was subsequently unavailable. **Other types of non-interviews** (codes 2.3, O) include persons currently ill (3) and cases of language barriers (65).

Partial interviews (56 cases, code 1.2, P) were kept as part of the the net sample, when at least 50% (and up to 80%) of all applicable questions were answered with a valid answer. Interviews with at least 80% valid answers count as full interviews (4,891 cases, code 1.1, I).

Table 1. Ghana mobile phone statistics, sampling frame, gross and net phone survey sample and weights

I. Mobile phone market statistics	Vodafone	MTN	Airtel-Tigo	Glo	TOTAL
Active subscriptions, voice calls (Oct 2022)[1)]	7,579,598	26,925,691	6,110,767	308,553	**40,924,609**

II. Sampling frame, gross sample	Vodafone	MTN	Airtel-Tigo	Glo	TOTAL
Overall gross sample	1,915	6,665	2,310	110	**11,000**

III. Worked gross sample		Vodafone	MTN	Airtel-Tigo	Glo	TOTAL
Interviews	(1.0)	753	3,277	891	26	4,947
Complete interviews	(1.1, I)	745	3,238	882	26	4,891
Partial interviews	(1.2, P)	8	39	9	0	56
Non-interviews, eligible	(2.0)	142	719	171	8	1,040
Refusal	(2.11, R)	100	507	104	6	717
Break-off	(2.12, R)	0	6	1	0	7
Non-contact, not available (2.2, NC)		37	163	47	1	248
No interview, illness	(2.32, O)	0	1	2	0	3
No interview, language	(2.33, O)	5	42	17	1	65
Unknown eligibility	(3.0, U)	358	900	468	26	1,752
No answer	(3.13)	358	900	468	26	1,752
Not eligible	(4.0)	195	314	180	24	713
Out of sample	(4.1)	14	22	7	1	44
Non-working/disconnected number (4.3)		164	212	140	23	539
Quota filled	(4.8)	17	80	33	0	130
Working numbers called[2]		1,284	4,998	1,570	61	7,913
Total numbers called		1,448	5,210	1,710	84	8,452

IV. Weights, contact, response and refusal rates	Vodafone	MTN	Airtel-Tigo	Glo	TOTAL
Probability of inclusion	.00017	.00019	.00026	.00020	-
Design weights, normalised	1.141	1.042	.753	.978	1
Combined weights, normalised	1.196	1.000	.813	1.672	1
Contact rates[3]	68.5%	78.3%	66.3%	55.0%	74.2%
Response rates[4]	60.1%	66.9%	58.2%	43.3%	63.9%
Refusal rates[5]	8.0%	10.5%	6.9%	10.0%	9.4%

1) NCA report, Telecom Voice Subscription Jan-Dec 2022
2) Sum of interviews, non-interviews, unknown eligibility & ineligible (but working numbers): I + P + R + NC + O + U + 4.1 + 4.8
3) The contact rate CON1 suggested by AAPOR: $(I + P + R + O)/(I + P + R + NC + O + U)$
4) The response rate RR2 suggested by AAPOR: $(I + P)/(I + P + R + NC + O + U)$
5) The refusal rate REF1 suggested by AAPOR: $R/(I + P + R + NC + O + U)$

Sources: Sampling data, gross sample, final survey data (v1.5)

The **overall contact rate** was 74% (using the contact rate CON1 suggested by AAPOR, see table 1 and footnote 3). The contact rate measures the share of effective contacts with respondents out of all eligible units.

The **overall response rate** was 64% (using the response rate RR2 suggested by AAPOR, see table 1 and footnote 4). The response rate is the share of interviews out of all eligible units.

The **overall refusal rate** was 9% (using the response rate RR2 suggested by AAPOR, see table 1 and footnote 5). The refusal rate is the share of refusals or break-offs out of all eligible units.

3.5. Questionnaire development and translations, programming

The questionnaire was developed in **English** and programmed in **SurveyCTO**, the software used for data collection.

3.6. Recruitment and training of interviewers

GSS recruited **60 interviewers** and organised a **three-day training** from 16 to 18 November 2022, held virtually via Zoom.

3.7. Pilot interviews

For the pilot interviews on 18 November, a sample of 421 numbers - taken from the gross sample - was called by interviewers. 167 interviews were completed, resulting in a total response rate of 41% (lower than for the full survey).

3.8. Data collection (incl. logistics)

The full data collection started on 25 November 2022. Out of the gross sample, 8,452 numbers were shared with interviewers and called. Interviews were conducted on all days of the week, although fewer on Sundays. Interviewers always used mobile phones, SIM cards and tablets provided by GSS. The phone credit for the calls was also provided by GSS.

Interviewers were instructed to attempt at least 5 calls to contact an individual. Non-Response was to be marked as final:

- after explicit refusal
- in case of non-eligibility of respondent
- after 2 unsuccessful call attempts: wrong number
- after 5 unsuccessful call attempts: if the number was currently not available or in case of a free signal, but no response

Instructions were also given to interviewers to always ask for the main user of the number/SIM card. If the main user was currently not available, a call-back time was to be noted and a call-back made.

As compensation for their time, respondents were offered an incentive in the form of phone credit, worth 10 Ghana Cedi (~ 1 USD). This credit was transferred to the respondent after the interview had been completed.

Data collection finished on 11 December 2022.

3.9. Accuracy and reliability

With a population size of around 20.7 million mobile phone users in Ghana (data from the 2021 census) and a sample size of n=4,947 the **standard error for proportion estimates is up to 0.7%** for the full sample. Region-level or district-level sampling errors depend on the subsample size of the respective area.

4. Quality management

4.1. Data cleaning

Extensive data quality checks were conducted in two phases, first by GSS, during and after data collection, and subsequently by Flowminder's survey team. Data cleaning included checks and edits for duplicates, ineligible mobile phone numbers, data entry errors, consistency errors, filter errors as well as outliers and highly implausible cases.

4.2. Weights

Two sets of weights are provided with the dataset. **Design weights** account for the sampling probability of respondents. For each of the Mobile Network Operators, sampling probabilities were calculated based on the number of calls placed to working numbers and the number of active phone numbers reported by the regulator (see table 1). Although the sampling was not stratified by MNO, differences in sample realisation of some MNO groups during data collection could have influenced the actual sampling probabilities. The sampling probability π_i of MNO group i can thus be written as π; = n_i/N_i, where n_i is the total of working numbers called from MNO group i, and Ni is the total of all existing working numbers of MNO group i. The design weights di are then the inverse of the sampling probabilities πi (di = πi/1, see Valliant et al. 2013: 311f). These weights were transformed to normalised design weights (mean=1). The design weights can be used for calculating user-generated weights.

The **combined weights** include the design weights and further adjustments for the MNO-specific non-response rates (see Valliant et al. 2013: 316f). When response rates differ between groups - here MNO groups - this additional selection effect might influence the relative weights of cases. The MNO-specific response rates were used to create non-response adjustment factors. The combined weights were then calculated as the product of the standardised design weights and the standardised non-response adjustment factors (see table 1 and Groves et al. 2009: 352), and also normalised (mean=1). These combined weights should be used for all inferential analyses.

5. Data management and archiving

Flowminder, as the project lead and as a research-driven organisation, will keep the data for future research purposes. Future requests for access to microdata (e.g. by academia) will be handled by Flowminder.

6. Citation

Flowminder Foundation and Ghana Statistical Services (2023). Ghana Mobile Phone Survey 2022. Dataset. Version 1.5.

Flowminder Foundation (2023). Methodology report for the Ghana Mobile Phone Survey 2022.

7. References

AAPOR (2023) Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 10th edition. The American Association for Public Opinion Research.

DHS Model Questionnaires - Phase 8. June 2020. https://dhsprogram.com/publications/publication-DHSQ8-DHS-Questionnaires-and-Manuals.cfm

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). Survey Methodology (Second Edi). Wiley.

Valliant, R., Dever, J. A., & Kreuter, F. (2013). Practical Tools for Designing and Weighting Survey Samples. Springer New York.