

FLOWMINDER.ORG

Methodology report | Phone survey

Democratic Republic of the Congo (DRC) Mobile Phone Survey 2021

Published: December 2024

Democratic Republic of the Congo (DRC)

Mobile Phone Survey 2021

Methodology report

1. Overall aim and purpose

1.1. Project context

The project **GRID3 Mapping for Health in DRC** (GRID3 M4H), implemented from 2020 to 2022, aimed to strengthen the effectiveness and equity of vaccination interventions in the Democratic Republic of the Congo (DRC) by using novel data and methods and through capacity building. Mapping for Health (M4H) in DRC was an initiative from the Ministry of Public Health, Hygiene and Prevention, **funded by GAVI, the Vaccine Alliance**, and implemented by Flowminder Foundation and the Center for International Earth Science Information Network (CIESIN) at Columbia University. The project was an extension of work conducted or supported in the DRC by the **Geo-Referenced Infrastructure and Demographic Data for Development (GRID3) programme**.

As part of this larger project, Flowminder Foundation was tasked with providing a range of technical assistance and data inputs. For Work Package 3 of the project (producing nationwide estimates of routine and sudden-onset population movements), a telephone survey among phone users in the DRC was designed, commissioned and analysed in order to gain empirical insights into phone use and mobility in the DRC. The objective of the phone survey 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 the DRC.

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.

The Khana Group (TKG, based in the US, Nigeria and DRC) was selected as the implementing organisation and was responsible for questionnaire translation, programming and testing, selection and training of interviewers as well as data collection in DRC, quality monitoring and initial data collection.

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

The **ethics committee of the Kinshasa School of Public Health** (KSPH) reviewed and approved the survey design and questionnaires.

From the outset of the activities, the DRC Statistics Institute, **l'Institut National de la Statistique (INS)**, was regularly informed about the progress of the survey. The INS also issued a visa (approval) for data collection. The final dataset and the methodological documentation have been shared with the INS for further statistical analysis and archiving purposes.

1.3. Content and topics

Current data about population characteristics and an in-depth knowledge of it are vital for health planning, risk management, disaster response, relief and humanitarian work, evaluation studies, demographic and social research, social statistics and many other use cases. The aim of the GRID3 M4H telephone survey was to understand more about the mobility of the DRC 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 DRC.

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 (e.g. users with different use frequencies, or users of different mobile network providers)
- 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 Z - Interviewer questionnaire

2. Legal basis and data protection

The initial lawful basis for data processing was the performance of a task carried out in the public interest, see GDPR Article 6(1)(e). The funding contract with GAVI outlines this task:

"[Flowminder] will support the Resource Mobilisation & Private Sector Partnerships Team in providing the Government of DRC with access to essential data on settlements and population distributions, characteristics and mobility patterns, as well as analysis of underlying gender-related barriers affecting vaccination coverage. Combining these datasets will enable effective analysis of access to vaccination services and support improvements to planning for wider coverage."

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, outlined the

purpose and goals of the survey, the project context, including the funder (GAVI) and the commissioning organisation (Flowminder), 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 the National Statistics Institute (INS). In that way informed consent was secured.

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/sondage.

3. Methodology

3.1. Mode of data collection

Computer-Assisted Telephone Interviews (CATI), i.e. individual telephone interviews with DRC phone users.

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 health planning purposes - and Flowminder's related analyses needs.

The indicators in **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 microcensus, which was also designed and coordinated by Flowminder in cooperation with WorldPop at the University of Southampton and the Kinshasa School for Public Health.

For the classification of attempted, partially completed and completed codes, the **disposition codes** suggested by AAPOR (2023) were used.

3.3. Statistical population

Primary Sampling Units were individual phone users who (a) used a SIM with a phone number from a DRC Mobile Network Operator, (b) were the main user of that SIM card, (c) were resident in the DRC at the time of the call, and (d) were at least 15 years of age.

Geographic coverage: The data include respondents in all 26 DRC provinces.

Reference period: Interviews were conducted between 15 October and 8 November 2021.

3.4. Sampling and sample size

The survey was designed as a cross-sectional **simple random sample (SRS)** telephone survey. No stratification was applied.

The target population was defined as all mobile phone users in DRC who were using a phone number of a DRC Mobile Network Operator (MNO). The existing network prefixes at the time of data collection were:

Vodacom : +24381, +24382, +24383
Africell : +24390
Airtel: +24399, +24398, +24397
Orange : +24389, +24385, +24380, +24384

Random Digit Dialling (RDD) was used to create a sample of DRC mobile phone numbers. For this purpose, SampleSolutions was subcontracted by TKG as a technical partner. SampleSolutions provided a **gross sample of 165,000 DRC 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 (see table 1). 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. As a first step, this pre-screening sample was randomly selected from the potential sampling frame, resulting in an initial gross sample of 637,146 phone numbers. In a second step, these numbers were screened via a 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). This second step resulted in a sample of 208,643 existing and active numbers. Out of these numbers, a random subset of 165,000 numbers was selected and forwarded to TKG for interviewing. In a third step, Table 1 shows the totals and MNO-specific counts for each of these steps.

From this overall gross sample of 165,000 DRC cell phone numbers, 14,500 numbers were shared with interviewers and **14,120 numbers were called**. In summary, at the end of data collection and after data cleaning, a total of 13,325 interview attempts and a total of 12,848 called working numbers were recorded. The net sample consists of **7,523 interviews** (see table 1).

The two main reasons for unsuccessful calls were unanswered phone calls (2,482) and refusals (2,176). 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 U). Only in cases where there was an explicit message that the number dialled was incorrect, not assigned or not in service was that call attempt marked as a "non-working number" (477 calls). **Refusals** (code R) comprised 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** (also code R).

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

I. Mobile phone market statistics		Vodacom	Orange	Airtel	Africell	TOTAL
Active mobile subscriptions, prev. 90 days (June 2021) ¹⁾		15,536,446	12,676,096	12,517,987	4,124,999	44,855,528
II. Sampling frame, gross sample		Vodacom	Orange	Airtel	Africell	TOTAL
Overall gross sample		42,281	64,347	51,962	6,410	165,000
III. Worked gross sample	Code	Vodacom	Orange	Airtel	Africell	TOTAL
Interviews	1.0	1,967	2,634	2,664	258	7,523
Complete interviews	1.1 (I)	1,905	2,546	2,584	251	7,286
Partial interviews	1.2 (P)	62	88	80	7	237
Non-interviews, eligible	2.0	726	1,037	909	96	2,768
Refusal	2.11 (R)	560	837	705	74	2,176
Break-off	2.12 (R)	55	69	82	5	211
Non-contact, not available	2.2 (NC)	91	112	102	15	320
No interview, other	2.3 (O)	12	10	15	0	37
No interview, language	2.33 (O)	8	9	5	2	24
Unknown eligibility	3.0 (U)	578	1,103	717	84	2,482
No answer	3.13	578	1,103	717	84	2,482
Not eligible	4.0	163	189	176	24	552
Out of sample	4.1	33	21	16	5	75
Non-working/disconnected number	4.3	130	168	160	19	477
Working numbers called ²⁾	-	3,304	4,795	4,306	443	12,848
Total numbers called	-	3,434	4,963	4,466	462	13,325
IV. Weights, contact, response and refusal rates		Vodacom	Orange	Airtel	Africell	TOTAL
Probability of inclusion		0.021%	0.038%	0.034%	0.011%	0.029%
Design weights, normalised		1.347	0.757	0.833	2.667	-
Combined weights, normalised		1.311	0.829	0.771	2.677	-
Contact rates ³⁾		78.9%	74.2%	80.4%	76.9%	77.6%
Response rates ⁴⁾		60.1%	55.2%	62.1%	58.9%	58.9%
Refusal rates ⁵⁾		18.8%	19.0%	16.6%	17.3%	17.2%

1) [ARPTC report O2 2021](#)

2) Sum of interviews, non-interviews, unknown eligibility cases, out-of-sample cases: I + P + R + NC + O + U + out

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 (v2.1)

A comparatively small number of calls (75) resulted in “out of sample” cases, i.e. persons who answered the call but were not part of the population of the study. This comprised mainly minors younger than 15 and persons residing outside the DRC. **Non-contacts** (code NC) included 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 (320), for example when call-backs were agreed but the respondent was subsequently unavailable. **Other types of non-interviews** (code O) included persons currently ill, in transit, or with bad connectivity (37) and cases of language barriers (24).

Partial interviews (237 cases, code 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 (7,286 cases, code I)

3.5. Questionnaire development and translations, programming

The source questionnaire was developed in English and subsequently translated to **French** and four common national languages - **Swahili, Tshiluba, Lingala and Kikongo**. All questionnaire versions were programmed in **SurveyCTO**, the software used for data collection. All 6 language versions were available to interviewers.

3.6. Recruitment and training of interviewers

TKG recruited **65 interviewers** in Kinshasa, DRC, and organised a **two-day training** on 11 and 12 October 2021 in Kinshasa.

3.7. Pilot interviews

For the pre-tests between 12 and 14 October, a sample of 617 numbers - taken from the gross sample - was called by interviewers. **90 interviews were completed**, resulting in a total response rate of 14.6% (lower than for the full survey).

3.8. Data collection and logistics

The full data collection **started** on **15 October 2021**. Out of the gross sample, 14,500 numbers were shared with interviewers in batches, and 14,120 numbers were called. Interviews were conducted on all days of the week, although fewer on Sundays. Interviewers mostly used their own mobile phones for calls, in some cases phones were provided by TKG. The phone credit for the calls was provided by TKG.

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.

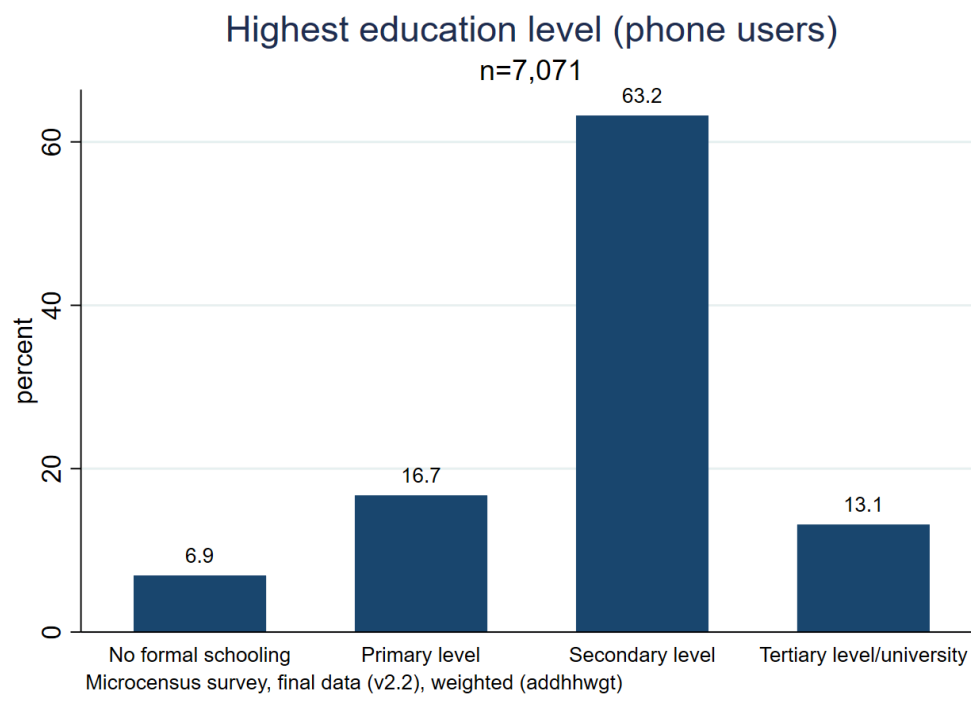
As compensation for their time, respondents were offered an incentive in the form of phone credit, worth 1 USD. This credit was transferred to the respondent after the interview had been completed.

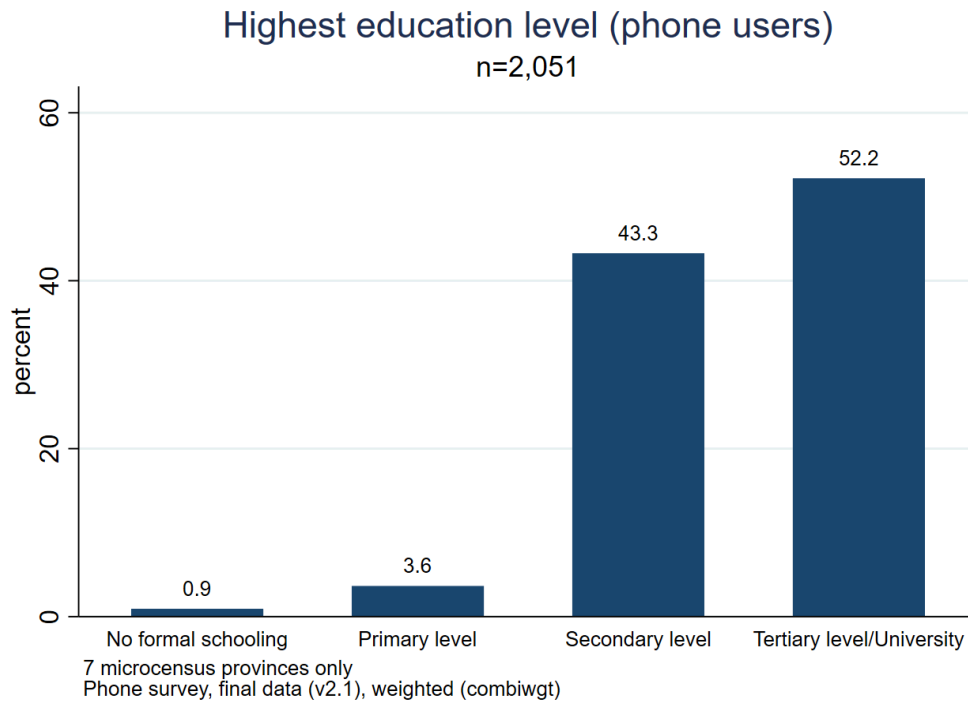
Data collection **finished** on **8 November 2021**.

3.9. Accuracy and reliability

With an estimated population size of around 45 million DRC mobile phone users and a sample size of **n=7,523**, the **standard error for proportion estimates is up to 0.6%** for the full sample. Province-level standard errors depend on the subsample size of the respective province.

Comparative data analysis between this phone survey (Oct/Nov 2021) and the microcensus implemented in seven DRC provinces earlier that same year (Mar/April 2021) point to substantial selection biases in the phone survey data. Based on the microcensus data, around 13% of mobile phone users held a tertiary educational degree. In the phone survey data, the share of that group in the microcensus provinces is much higher, at 52% (also 52% across all provinces). Correspondingly, phone users with no formal schooling or primary education seem to be underrepresented in the phone survey data. A similar effect can be observed for household-level socio-economic status - an overrepresentation of high-status individuals in the phone survey.





4. Quality management

4.1. Data cleaning

Extensive data quality checks were conducted in two phases, first by TKG, 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. Complete interviews with a duration of less than 5 minutes - pointing to quality problems - were dropped from the dataset. All string variables were checked and coded to categorical variables where applicable.

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 MNO groups during data collection influenced the actual sampling probabilities. The sampling probability π_i of MNO group i can thus be written as $\pi_i = n_i/N_i$, where n_i is the total of working numbers called from MNO group i , and N_i is the total of all existing working numbers of MNO group i . The design weights d_i are then the inverse of the sampling probabilities π_i ($d_i = 1/\pi_i$, 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, which were also normalised (mean=1). The combined weights were then calculated as the product of the normalised design weights and the normalised non-response adjustment factors (see table 1 and Groves et al. 2009: 352). These combined weights should be used for all inferential analyses.

5. Data management and archiving

After the end of the project, the data and metadata will be shared with the DRC INS for archiving and other statistical purposes. This is in line with the principle of national ownership of data collected, as specified by the donor contract. Flowminder, as the project lead and as a research-driven organisation, will keep the data for future research purposes as well. Both the INS and Flowminder will have ownership of the data after the end of the project. Future requests for access to microdata (e.g. by academia) will be handled by one of the two organisations.

6. Citation

Flowminder Foundation (2022). DRC Mobile Phone Survey 2021. Dataset. Version 2.2.

Flowminder Foundation (2022). Methodology report for the DRC Mobile Phone Survey 2021.

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 Edn). Wiley.

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