FLOWMINDER.ORG

Correcting for biases in mobility indicators derived from Call Detail Records

UN Big Data Regional Hub for Africa

21 March 2024

Zachary Strain-Fajth
Data Analyst
Flowminder Foundation

Outline

- Why
  - Inferences from CDR data
  - Differences in phone use
  - Differences in mobility
- How
  - Proposed method for bias adjustment & scaling
  - Formula
  - Data sources
- What
  - Country examples
  - Next steps

The promise of Big Data

Using mobile network operator (MNO) data for the public good

MNO source data

`MSISDN | DATETIME | CELLID`

e.g. Jan 2020 - Dec 2023

Cell geolocation

Compared to survey and census data, MNO data have **several advantages**:

- Automated data collection by MNOs for billing purposes
- No primary data collection required
- Once access to data is set up, low maintenance costs
- **Very high temporal granularity**
- **High spatial granularity**
- **Recency/timeliness:** data become available within a few days

But also **disadvantages**

- Setting up partnerships can be a lengthy process
- High set-up costs
- Lack of rigorous validation data
- **How to quantify uncertainty?**
- **Selectivity and biases of MNO data**

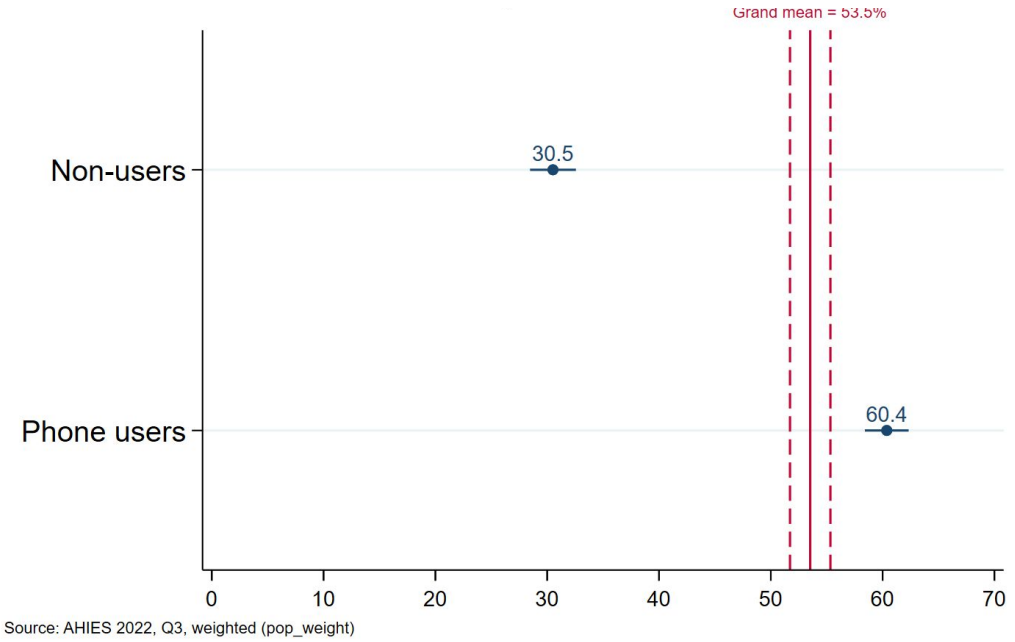
Biases due to mobility differentials

Firstly, **mobility is often different** between mobile phone users and **non-users**

- With mobile phone users showing **higher mobility**, on average, **than non-users**
- Mobility estimates based on phone user data alone often **overestimate** (at times: underestimate) **mobility**
- Not all mobility indicators, however, show such large differences between these two groups

Ghana: % of population who did a 5+km trip in past 3 months

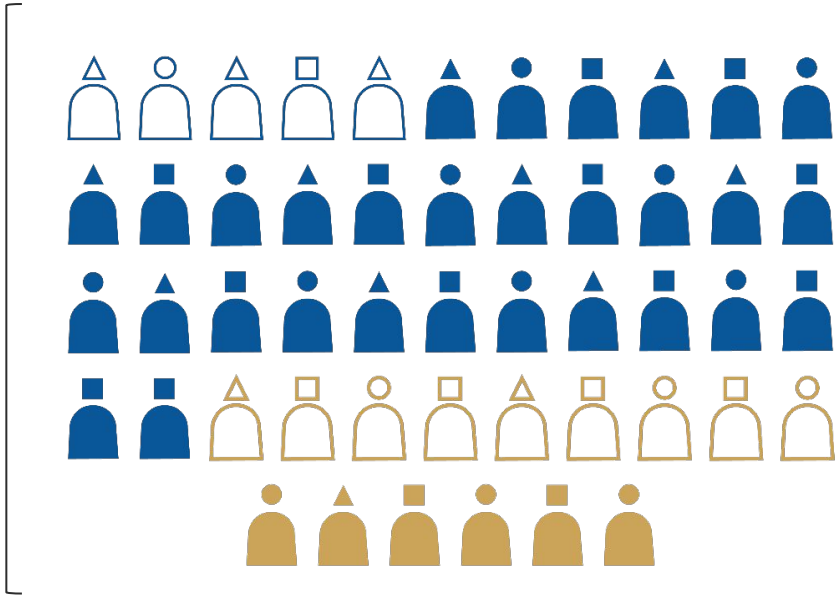
n=30,928



Source: AHIES 2022, Q3, weighted (pop_weight)

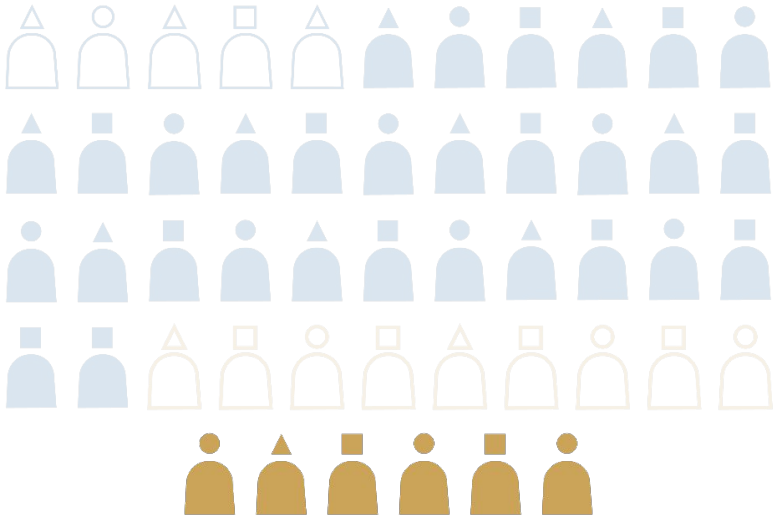
Not-so-big data: the issue of bias

National (adult) population



- No mobile subscription
- Subscribers to other mobile network operators (MNOs)
- "Inactive" subscribers of participating MNO
- **"Active" subscribers of participating MNO**

Not-so-big data: the issue of bias

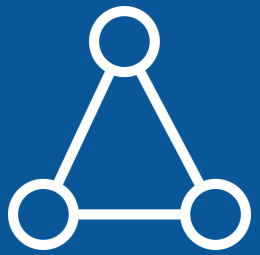


- No mobile subscription
- Subscribers to other mobile network operators (MNOs)
- "Inactive" subscribers of participating MNO
- **"Active" subscribers of participating MNO**

* Non-random 'sample' available for analysis

Additional data on demographics, phone use and mobility from field and phone surveys can help to address biases and therefore to get the most out of CDR data.

MNO subscribers are not a random sample of the population, nor can be assumed to be.



The need for data triangulation (data fusion)

CDR aggregates alone are not sufficient for

- Estimates on **population counts** (stocks) or **population density**
- Estimates on **population change** (births/deaths, immigration/emigration) or population density change
- **Extrapolation outside coverage areas** of CDR data

Survey (and census) data can and should be used!

Flowminder has recently developed estimation methods to adjust for representation biases & provide population-scaled estimates for

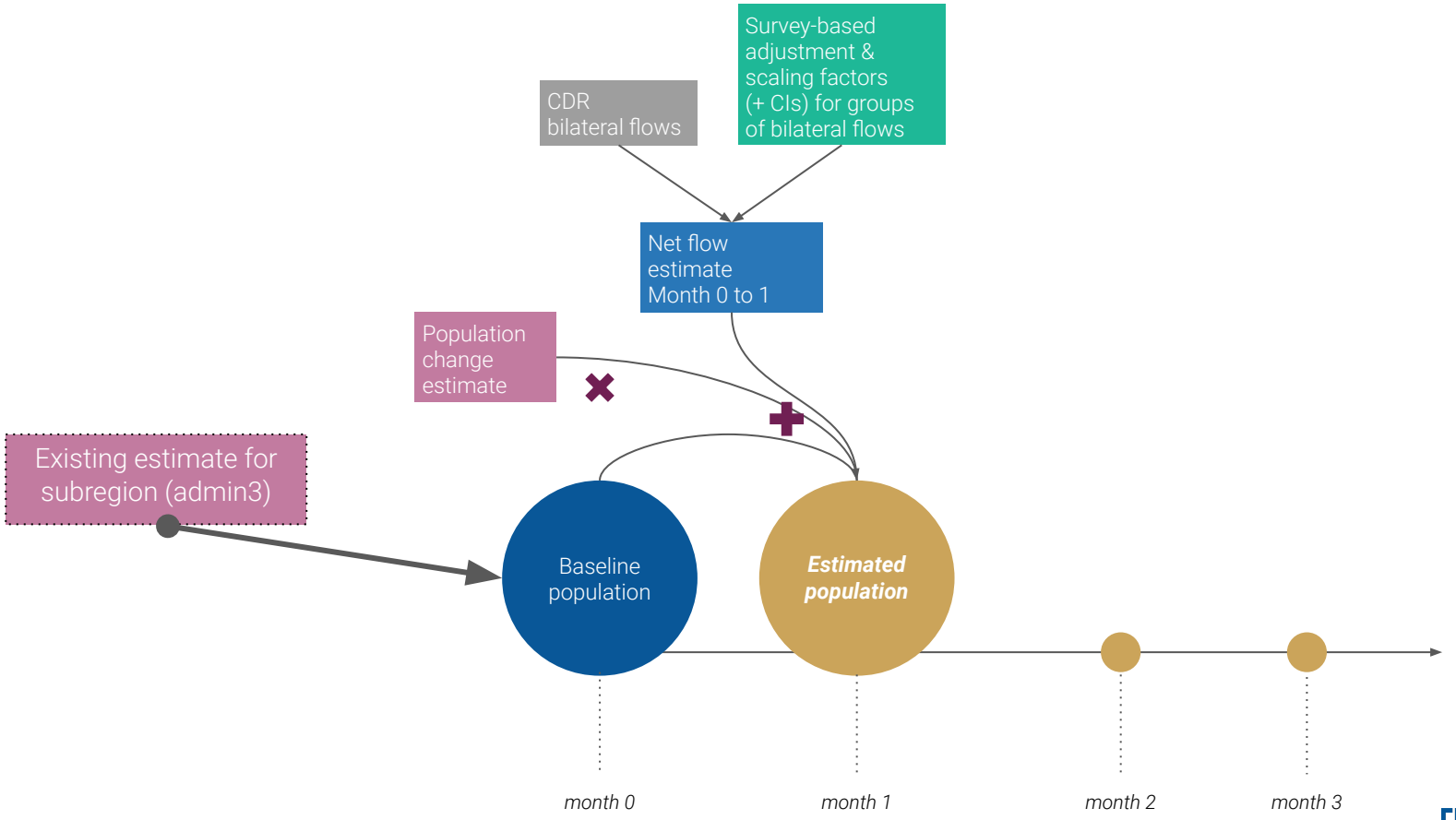
- **Relocations** from sub-region to sub-region, per month
- **Residents** per sub-region, per month

Bias-adjusted and population-scaled (weighted) estimates

These estimates are based on

- CDR aggregates
- Primary & secondary survey data
- Existing population estimates

Method for monthly residents' estimates



Method for relocations estimates

- Relocations from area a to area b between month m and month n can be estimated from CDR aggregates of relocations (cdr_flow_{abmn}) between those areas and months, and from a flow **adjustment factor** and a flow **scaling factor**
- Flows are adjusted for the number of SIMs per user ($sims_{ab}$). The flow scaling factor is the inverse of the share of MNO users (mno_share_{ab}) in the flows:

$$est\_flow_{abmn} = cdr\_flow_{abmn} * (1/sims_{ab}) * (1/mno\_share_{ab})$$

- Note: Parameters for the subset of mobile households/individuals only available at admin1 by admin1 level

Method for monthly residents' estimates

- The estimate of residents in area a for month n ($est\_residents_{an}$) is calculated as the sum of the baseline population for that area ($est\_base\_pop_{a}$) and by iteratively adding the cumulative sum of all net arrivals ($est\_netflow_{amn}$) for all months between the baseline month and the current month, and by applying an area-specific rate of natural population growth ($growthrate_{a}$) to each monthly sum:

$$est\_residents_{a1} = est\_base\_pop_{a} \quad \text{(Month 1 (baseline), m=0, n=1)}$$

$$est\_residents_{a2} = (est\_residents_{a1} + est\_netflow_{a12}) * changerate_{a} \quad \text{(Month 2, m=1, n=2)}$$

$$est\_residents_{a3} = (est\_residents_{a2} + est\_netflow_{a23}) * changerate_{a} \quad \text{(Month 3, m=2, n=3)}$$

$$\ldots = \ldots$$

$$est\_residents_{an} = (est\_residents_{am} + est\_netflow_{amn}) * changerate_{a}$$

- where the net flow estimate for area a between months m and n is the sum of all estimated inflows to that area minus all estimated outflows from that area:

$$est\_netflow_{amn} = est\_inflow_{amn} - est\_outflow_{amn}$$

Data sources

Democratic Republic of the Congo

WorldPop 2020: gridded population estimates

Micro-census 2021: covering 7 provinces, phone users and non-users (led by FM)

Phone survey 2021, targeting phone users across the country from all MNOs (commissioned by FM)

Ghana

Annual Household Income and Expenditures Survey (AHIES) 2022: phone users and non-users (GSS)

Census 2021: phone users and non-users, population estimates (GSS)

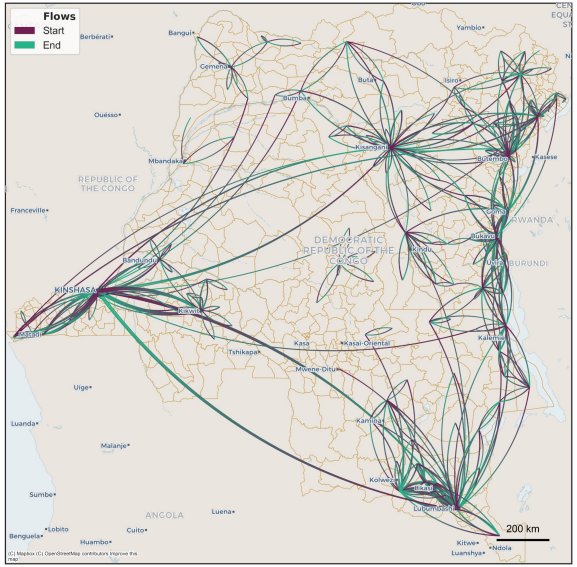
Phone survey 2022: targeting phone users across the country from all MNOs (commissioned by FM, conducted by GSS)

Producing monthly mobility & population estimates

FLOWMINDER.ORG

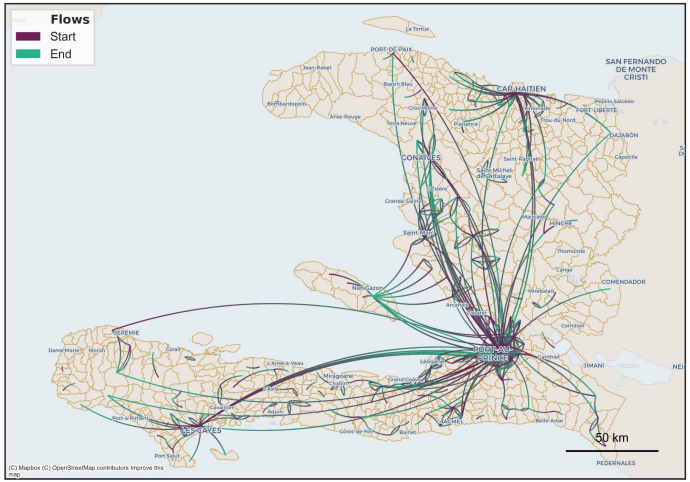
Monthly relocations between sub-regions

DRC



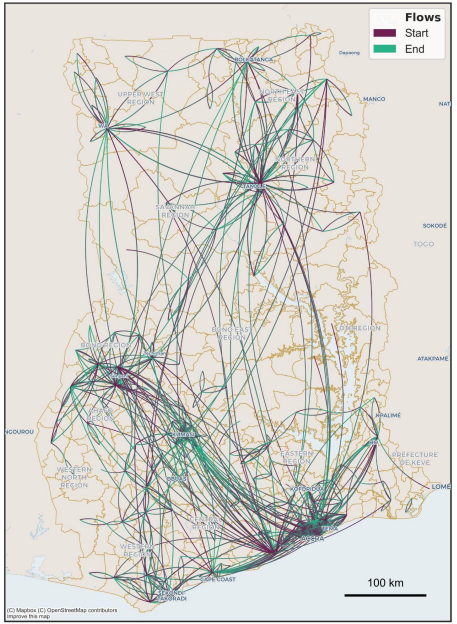
Note: estimated top 1,000 flows between health zones, median, Nov 2021 - Dec 2022

Haiti



Note: top 500 flows between communal sections, median, Feb 2020 - Feb 2022

Ghana



Note: unscaled top 1,000 flows between districts, median, Jan - July 2021

FLOWMINDER.ORG

Next steps

Ensure regular updates to parameters (i.e. pop. change %, mobility, phone use) and **onboarding additional MNOs**

Refinement of adjustment and scaling factors (in progress)

Testing estimations with **validation** data

Paper forthcoming!

Thank you!

FLOWMINDER.ORG

FLOWMINDER.ORG

Zachary Strain-Fajth
Data Analyst
zachary.strain@flowminder.org

 www.flowminder.org

 info@flowminder.org

 @Flowminder