

Building a Survey without a Recent Census

The Case of the Sudan Labor Market Panel Survey 2022

Caroline Krafft (St. Catherine University)

Ragui Assaad (University of Minnesota)

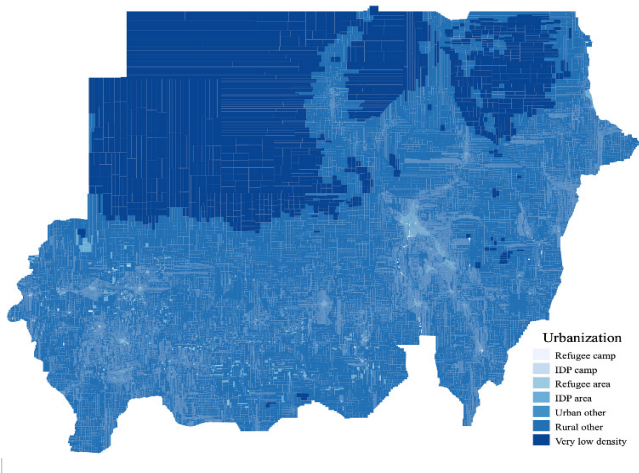


How can countries without a recent census still collect survey data? This brief explores how geospatial data filled the gap in Sudan.

Follow us on Twitter! Find @glmlc

Topic at a Glance

Low-income countries are also often data poor, lacking recent censuses or surveys to inform policy and programs. Without a recent national population census, countries do not know how to sample for nationally-representative surveys. This brief discusses how countries can generate nationally-representative surveys without a recent census by relying on publicly available high resolution geospatial data. The brief discusses how such geospatial data and sampling were used to create the Sudan Labor Market Panel Survey (SLMPS) 2022. Sudan's last population census was in 2008, a number of years ago, and before the country was divided. The SLMPS is Sudan's first nationally representative household survey since 2014 and first labor force survey since 2011. These methods can be used in other contexts that lack a recent official census or for sampling between censuses in areas with substantial changes in population.



Caption: Sudan full sample of primary sampling units (PSUs), classified by area type. Source: See details in Krafft, Assaad, and Cheung (2023).

New Insights

Low-income and especially fragile and conflict-affected countries are often data-deprived (Corral et al. 2020; Wardrop et al. 2018). The Middle East and North Africa (MENA) region is particularly short on both microdata and economics research (Das et al. 2013; Ekhatior-Mobayode and Hoogeveen 2021). Sudan is one such low-income, fragile, and conflict-affected country in MENA. The last population census in Sudan was in 2008, before the South seceded (Minnesota Population Center 2020). Sudan's last labor force survey was in 2011 and its last nationally-representative household survey of any kind was in 2014 (Ebaidalla and Nour 2021).

In order to fill this data gap, with funding support from G2LM | LIC and the World Bank, the Economic Research Forum (ERF) and Sudan's Central Bureau of Statistics (CBS), in collaboration with the Ministry of Labor, undertook the Sudan Labor Market Panel Survey (SLMPS) 2022. The LMPSS, which have also been fielded in Egypt, Jordan, and Tunisia (Krafft, Assaad, and Rahman 2021; Krafft and Assaad 2021; Assaad et al. 2016), are a multi-purpose household survey with a strong focus not only on labor but also on gender.

The project faced an initial challenge in that the population census was outdated. Moreover, while there were defined and updated boundaries for the first level of administrative geography (the state, N=18) and the second level of administrative geography (the municipality, N=189), lower levels of geographic boundaries and primary sampling units (PSUs, clusters for sampling households within) had not been updated or generated.

In order to generate a nationally representative sample, the SLMPS 2022 instead relied on geospatial data. Specifically, the SLMPS 2022 used WorldPop geospatial files with 100 meter x 100 meter population estimates (WorldPop 2020). We used an algorithm to combine pixels into PSUs with an estimate of around 50-100 households, but with tractable (rectangular) boundaries and within the same municipality for feasibility of fielding.

Each PSU was classified as a refugee camp, refugee area, IDP camp, IDP area, urban (other), rural (other) or very low density (less than 0.025 persons per square kilometer, excluded from sampling). The figure shows the full sample frame of PSUs, classified by area type. We therefore could undertake stratified, random sampling, probability proportional to size (population). Within each PSU, households were systematically listed and 20 households per PSU were sampled. The strata allowed the SLMPS to over-sample refugees and IDPs to ensure an analyzable sample of these vulnerable groups. The population estimates and sampling strategy allowed for sample weights to be generated to ensure nationally representative statistics even with over-sampling. See Krafft, Assaad, and Cheung (2023) for details on the SLMPS 2022.

The SLMPS 2022 data will be publicly available starting at the end of August 2023 from ERF (at www.erfdataportal.com). They will serve as an important, nationally-representative data public good. Research is already underway on important topics such as gender and labor force participation and women's economic empowerment.

Policy Recommendations

Technical assistance is needed for national statistics offices to implement geospatial sampling methods. For the SLMPS 2022:

- The project principal investigators developed the Stata (statistical software) and QGIS (geographic information system software) code and protocols.
- The principal investigators and ERF then trained CBS on how to implement the code and protocols, with both conceptual training on sampling and hands-on training replicating the SLMPS 2022 Stata and QGIS sample frame creation and sampling.
- CBS is now able to use the sample frame to create other nationally-representative samples for future surveys.

International support is also needed for low-income countries to be able to undertake more, and more frequent, population censuses and nationally-representative surveys.

- Sudan's CBS had no budget to implement surveys or censuses on its own, without the funding of G²LM | LIC, and other low-income countries face similar constraints.
- Geospatial sample frames and samples are a viable, but second-best, alternative to population censuses and nationally representative surveys based on sample frames from recent population censuses.

- Financial and technical support are both important inputs to creating population censuses and nationally-representative surveys.

- Such censuses and surveys should be made publicly available to serve as data public goods for research, policy, and programming. Random sub-samples (e.g. a 10% sample of a census or 50% sample of a survey) are one option if countries do not want to release the full data.

Limitations

Using geospatial data to estimate populations, create sample frames, and undertake nationally-representative surveys is an important option in the absence of a recent population census to serve as a nationally-representative sampling frame. It can also be helpful in areas that have had rapid population changes between regular censuses. This approach is, however, less accurate than a true population census. It should only be considered as an interim measure, not a substitute for undertaking a population census. Moreover, much more research is needed to understand the degree of random or systematic error when using geospatial data to sample.

Read more → g2lm-lic.iza.org/projects/ta3/advancing-data-capacity-for-policy-innovation-in-sudan-labor-market-panel-survey-2019/

G²LM|LIC is a joint program by IZA and the UK Foreign, Commonwealth & Development Office (FCDO) – improving worldwide knowledge on labor market issues in low-income countries (LICs) and providing a solid basis for capacity building and development of future labor market policies.