

RESEARCH BRIEF

Monitoring the SDGs by means of the census

**RESEARCH LEAD: TOM A MOULTRIE,
UNIVERSITY OF CAPE TOWN - CENTRE FOR ACTUARIAL RESEARCH**

TRENDS

Thematic Research Network
on Data and Statistics

CONCEPT

DEFINING IDEAS

- Population-related elements are present in approximately 40% of the SDG indicators. Without appropriate, accurate, and timely data, framed by the principle of 'no-one left behind', our ability to monitor progress towards meeting the Sustainable Development Goals will be constrained.
- The challenges of collecting and employing such data will be borne, largely, by the countries of the global South.

The Sustainable Development Goals (SDGs), agreed to by the UN in 2015, pose a significant and fundamental challenge to the global development community, and will shape the global development agenda over the period to 2030. The 17 goals, with their associated 169 targets and more than 230 unique indicators, are highly ambitious and meeting them will require a monumental sustained effort from nations, governments, civil society organizations, as well as the general populace. Achieving the SDGs by 2030 will also require that progress towards meeting the goals be monitored and tracked on a regular and systematic basis, which in turn will present substantial data challenges that require the urgent attention of the global data community.

While the SDGs cover all areas of development including mitigating climate change, and moving towards a more sustainable ecosystem, population-related data will play a hugely important role in assessing our progress towards meeting the SDGs: in the official list of indicators, 97 (some 40 per cent) have a population-related element in either numerator or denominator, or both. Without appropriate means of measuring these population-related elements of the indicators, our ability to accurately track progress will be severely constrained.

The focus of this research brief is on these population-related indicators. Where might they be sourced from? What data challenges do they present? Within this focus, our particular attention is on the countries of the global South. For the countries in the global North, the data required to monitor and track the SDGs are mostly readily accessible and the data challenges are not insurmountable. In these countries, many indicators will be able to

be monitored in near-real time. In the global South, however, the issues are somewhat more intractable. Administrative databases, such as civil and vital registration systems or population, taxation, or schooling registers are frequently substantially incomplete and generally not fit for the purpose of tracking the SDGs.

For many countries in the global South, the census is a national data collection exercise without equal in terms of scale and complexity. In the absence of substantially complete vital registration or administrative databases, the census provides the only indication of those most basic features of a country's population. The headline population counts provide a snapshot of the size and distribution of a country's population at the census date, even while the processing, cleaning and preparation of tabulations from the data collected in the census take several years after the data are collected.

The universal scope of the census, that is its attempt to enumerate all people in a defined area at a point in time, also means that the census is frequently the only potentially reliable source of information on minority populations, on those at risk of being left behind, as well as providing data on rare or hard to measure events, such as migration and mortality.

For these countries, their population census is a vitally important source of data that will be called upon directly for monitoring progress towards meeting the SDGs, and which will form the basis for a number of other data collection exercises that will be required to produce the highly granular and spatially disaggregated data that will be required to ensure that no-one is left behind.

This policy brief highlights some of the important implications of countries having to rely, directly or indirectly, on their census data in the SDG process, and draws attention to possible interventions and strategies that could be used to mitigate these implications.

The timing of the census in most of the global South means that censuses in the 2020 round have already occurred, or will occur too late to baseline the SDG goals. Furthermore, there is going to be a need to update the census data on an ongoing basis to ensure that the indicators are timely and relevant.

MAKING SENSE OF THE CENSUS

The United Nations Statistics Division (UNSD) plays a central role in guiding and capacitating the conduct of censuses around the world. Not only does UNSD offer technical advice on the design, implementation, conduct, collection and analysis of census data, they also make recommendations as to what information should be collected in censuses.

Through the UNSD, the United Nations recommends that all countries conduct a census at least every ten years, preferably in years ending in zero or one. Each ‘round’ of censuses covers a period of ten years, with the ‘2020 round’ covering all censuses conducted between 2015 and 2024. A few developed countries (e.g. Australia and Canada), and several small/island states run censuses more frequently. Moreover, a number of countries in the global North do not run censuses directly, preferring to conduct virtual censuses that allow a detailed picture of their populations to be established from the population, tax, residence, schooling, and other administrative databases that are constantly maintained and updated in real time.

Indeed, with high-quality administrative data capabilities, the census is largely redundant – although some countries that possess that capability still require a regular census to allocate representation in spheres of government, as is the case in the United States. Yet, the United States Bureau of the Census’ observation that the conduct of their decennial census is “the federal government’s largest and most complex peacetime operation” points to an awful irony: those countries where administrative data systems are weak and incomplete are also those most reliant on an exercise of immense logistical complexity, and those same countries frequently lack the human and other

resources to conduct such a complex exercise successfully. Almost without exception, this describes the challenge of collecting and analyzing population data in the global South.

In addition to the value of the information collected directly in a census, the census has important indirect significance too: samples for national and global survey programs (such as the Demographic and Health Surveys) are usually drawn from the results of the census. Where the census suffers systematic error, the quality of these data sources may be compromised.

POPULATION DATA FOR MONITORING THE SDGS

In short, in the global South particularly, the census will be the bedrock of a national statistical framework for monitoring the SDG agenda. But there are two aspects that should give cause for concern.

The first relates to the baselining of the SDGs. Despite having been launched over two years ago, few countries (and even fewer in the global South) have constructed a baseline of the SDG indicators at the start of the SDG period. Knowing how far we have progressed will be almost impossible if we do not know whence we started. The second relates to the ongoing effort required to monitor and track progress towards meeting the SDGs. Both require population data, which — in the case of the countries in the global South — will come largely from the census.

However, given that most countries in the global South conduct censuses only every 10 years, most countries will conduct either one or two censuses in the period covered by the SDG agenda, and for many of them it will be in the period 2020-1, with the second being conducted at the very end of the SDG period. For the 2020 round, the timing of censuses by region and year is shown in Table 1.

Nearly half of all African countries' censuses in the 2020 round will have occurred by 2019, although even among the 10 African where those censuses are scheduled for 2018 or 2019, the preparations for these censuses are probably too far advanced for significant interventions to be made to incorporate new questions in their questionnaires that would allow baselining or monitoring of the SDG agenda. And, where the next census is still some way off, the census will not be able to provide useful data against which the SDG indicators can be baselined.

Second, one needs to consider the fitness for purpose of census data for monitoring and tracking progress towards meeting the SDGs. In South Africa, the last three censuses (in 1996, 2001 and 2011) have undercounted the population by around 10, 17 and 15 percent respectively. While in itself cause for concern, the South African situation is perhaps better than in many other countries in the global South. Other countries may indeed enumerate their populations more

completely, yet the reason why we know that the South African censuses missed those proportions of the population is because a Post-Enumeration Survey (PES) was conducted in that country. A PES is, in effect, a capture-recapture exercise: running a much smaller survey in the weeks following the census to evaluate, on a sample basis, to validate the census enumeration. On account of its relatively small sample size, the PES is usually stratified on only a few variables: coarse groupings of age; sex; location, for example. From this survey, it is mathematically simple (albeit with the strong assumption that probability of being 'found' in each exercise is independent) to estimate the undercount in the census.

However, owing to the logistical complexity of running a further survey on the coattails of a census, the UNSD does not strongly advocate for countries to routinely conduct a PES. Without a PES, though, the true extent of who might have been missed in the census remains unknown.

But, even for those countries that do conduct a PES, in order to precisely understand who might not have been enumerated, and – perhaps, therefore – left behind in the measurement of SDG indicators, would require a sample size orders of magnitude larger (with commensurate increases in costs and logistical complexity) than those currently being conducted.

Table 1: Number of censuses by region per year between now and 2030

| | Africa | North America | South America | Asia | Europe | Oceania |
|----------------|--------|---------------|---------------|------|--------|---------|
| 2015 | 3 | 5 | 1 | 6 | 2 | 3 |
| 2016 | 7 | 2 | 0 | 5 | 3 | 5 |
| 2017 | 8 | 3 | 4 | 1 | 0 | 2 |
| 2018 | 5 | 0 | 0 | 3 | 1 | 2 |
| 2019 | 5 | 0 | 0 | 4 | 2 | 4 |
| 2020 | 6 | 14 | 3 | 16 | 4 | 5 |
| 2021 | 4 | 14 | 1 | 9 | 35 | 3 |
| 2022 | 4 | 2 | 4 | 2 | 1 | 2 |
| 2023 | 5 | 1 | 0 | 0 | 1 | 0 |
| 2024 | 6 | 0 | 0 | 2 | 1 | 0 |
| Unknown | 5 | 1 | 1 | 6 | 1 | 0 |
| TOTAL | 58 | 42 | 14 | 54 | 51 | 26 |
| 2015-19 | 28 | 10 | 5 | 19 | 8 | 16 |
| 2020-24 | 30 | 32 | 9 | 35 | 43 | 10 |

In summary, the census contains the raw data for providing many of elements required to measure the SDG indicators. Population counts, broken down into fine granularities and for small areas, can only be obtained through the census. However, the census remains imperfect. Most countries will have to rely on projection models to provide the baseline population counts required. The accuracy of the census in many countries in the global South remains questionable. Yet, without the census, measurement will prove near impossible.

Some elements of the 98 population-related indicators may be determined from functional and properly maintained and updated administrative systems. More indicator elements could be derived from surveys, vital registration systems, or population projections, thus reducing reliance on the census. Yet each of these have drawbacks, and each alternative is an imperfect substitute for another.

Surveys are more cost-effective (in aggregate, if not per capita) than censuses, and allow much greater detail to be collected on a limited number of topics from a smaller universe of respondents. Still, the greatest weakness associated with reliance on surveys is that, in order to ensure that the goal of no one left behind is achieved, the scale and sample size of surveys will have to be increased dramatically. This will come with increased cost. In addition, as mentioned above, since many surveys are sampled from a frame derived from a census, if the census is deficient, the sample universe may suffer from structural and systematic biases.

Administrative data systems are largely incomplete across the global South. An assessment from the UNSD suggests that in almost no developing country are civil and vital registration systems complete; and this calls into question whether they can be used to provide policy-relevant information, or can be used reliably as elements in the measurement and monitoring of the SDGs.

Similarly, data from different administrative databases in developing countries are seldom completely interoperable: more often than not, they lack crucial information that would allow one database to be merged with another.

Given these difficulties, population projections will play a significant role in the baselining and monitoring of the SDGs. Since censuses are

infrequent and CRVS incomplete, indicators that are expressed per capita will be able to be measured in real-time if appropriate denominators are available. Population projections, and other models, may indeed be able to provide that information. The risks here are manifold. First, the models or population projections may be mis-specified or incorrect. Second, the uncertainty inherent in any model will increase the more the data are disaggregated. Third, there is a danger of model results being mistaken for empirical observations. Fourth, there is a problem insofar as that the leading and most commonly used models and projections are all derived from institutions headquartered in the North. This could result in a situation where data providers in the global South are not capacitated, but will become disempowered consumers of their own data.

POLICY IMPLICATIONS

For a great many countries in the global South, the census, and projections derived therefrom, will be indispensable for baselining and monitoring progress to meeting the SDGs. Yet a number of important policy implications can be drawn from the analysis above.

First, there is an urgent need for a globally coherent program to establish the baselines for measuring and monitoring the SDGs. Since we are already more than two years into the SDG window, establishing the retrospective data should be a global priority. Achieving this would require that countries conduct data audits, to establish what information is readily accessible to baseline as many SDG indicators as possible. Specific consideration should be given to how best to identify gaps in coverage — especially at the fine levels of granularity and spatial disaggregation required to establish who is at risk of being left behind.

Second, countries need to develop a data plan for monitoring the SDGs. For most countries, it may prove impossible to monitor every indicator on an annual basis; where this is so, a revolving schedule of data collection efforts might allow indicators to be tracked every few years. To improve global comparability, it may be desirable to create and validate a global SDG survey, implemented in countries where administrative and other data are absent or substantially incomplete.

Third, we should not lose sight of the challenge of collecting and reporting administrative data in near-real time. The importance of improving civil and vital registration systems, particularly, along with other administrative systems, cannot be underestimated, and is currently the focus of much attention. In the longer-term, our goal should be to move to the position enjoyed by many countries in the global North. The SDGs can be used as further leverage to capacitate administrative, civil and vital registration system enhancements.

This, however, is a long-term goal. In the shorter-term, there is a desperate need for seeking to enhance the evidence base for baselining and monitoring the SDGs. In this regard, where countries still have time to modify their census questionnaires, national statistical offices should urgently consider which indicators require data that can only be derived from census data, and take all necessary steps to include them. In general though, countries are unlikely to accelerate their census plans: we will have to work with the data available. Furthermore, countries should be urged to run a PES after their census. Inaccurate census data may compromise the utility of the data collected, as well as downstream activities, such as survey sampling. Determining the scale of a PES is an important consideration to allow as much detailed information as possible to be established as to who might be being left behind.

Finally, there is a pressing need to develop methods for integrating and harmonizing data from a multiplicity of sources. Censuses provide detailed information on an individual level; mobile telephony data can tell us who is connected to a network, and where; remote sensing (satellite) data can provide near real-time information on urbanization; land use; and population density. Each of these data sources on their own is inadequate to the challenges; yet integrating and harmonizing these data offers the hope of a resolution to many of the issues raised.

LINGERING QUESTIONS

Even though the necessity of improving the evidence base to better monitor and track the SDGs is clear, and the path for doing so is well understood, there are still many obstacles in the way. Significant financial and human capital resources will have to be invested in national statistical systems to establish the requisite capacity and data systems to monitor and track the SDGs.

One important step has already been taken. The Cape Town Global Action Plan endorsed at the UN World Data Forum in January 2017 laid out the strategic priorities to meet the data challenge of the SDGs — strengthening core statistical processes (including censuses and surveys) and integrating geospatial data more widely in statistical data; strengthening and modernizing national statistical systems; establishing the basis for data interoperability and harmonization; the need to build partnerships outside the established statistical data ecosystem; and the need to mobilize resources to finance and capacitate these priority areas.

Yet the question remains as to whether the political will exists to give meaningful content to this Action Plan, and to harness and integrate old and new data forms in the service of the SDG agenda. Paying serious attention, not lip-service, to the SDGs and the overarching philosophy and principles that underpin them, will require all stakeholders — and the national statistical offices, in particular — to adopt a whole new approach to data, to national statistical systems and — especially — to give heft to the idea of the ‘data revolution’.