

May 2017

From figures to facts: making sense of socio- economic surveys in the Democratic Republic of the Congo (DRC)

Over the last decades, both donor organizations and the national government have invested heavily in administering national-level surveys in the DRC. These survey results have been used in planning exercises (e.g. to elaborate the Poverty Reduction Strategy Papers) and in evaluation exercises (e.g. to decide on the country's debt relief), but the quality of these data is highly contentious as some basic demographic information on the DRC is simply unknown. This policy brief first presents the issue and illustrates its direct implications before we present a discussion of possible solutions, also for the short-run.

Wim Marivoet

IFPRI West and Central Africa Office
w.marivoet@cgiar.org

Tom De Herdt

IOB, University of Antwerp
tom.deherdt@uantwerpen.be

Institute of Development Policy and
Management
University of Antwerp
iob@uantwerpen.be
+32 3 265 57 70

uantwerp.be/iob

The issue

Policymakers are typically concerned with absolute numbers of poor, ill, undernourished people; where they are located; and how their situation has evolved over time. In many developing countries, where often no routine monitoring system exists, one typically relies on representative household surveys to find answers to these questions. According to Little, survey sampling is one of the most important contributions of statistics to science as “[i]t provides the remarkable ability to obtain useful inferences about large populations from modest samples, with measurable uncertainty” (2004, p. 546). The crux of the matter of representativity lies however in the weight associated to each sampling unit, which indicates the number of population units it is supposed to “represent”. Only with accurate weights is it possible to calculate any statistic from the sample as *if it were* derived from the whole population.

Unfortunately, this weighing trick does not seem to work very well for the Democratic Republic of the Congo (DRC).

The basic problem is that over the years the demographic structure of the country became largely unknown. **All 21st century survey samples have been defined in function of population projections based on the outdated and incompletely processed census of 1984.** Indeed, over the 33 years

which separate us from this last census, too much has happened, ranging from the end of the Zairian state in 1996, over the two consecutive wars that followed, to a gradual but still fragile economic and institutional recovery since 2003. How all this impacted on the Congolese demography, we do not know. To give an idea, death toll estimates of the war seem to range from 200,000 people (Lambert and Lohlé-Tart, 2008) to 4.5 million (Coghlan et al., 2008). And for migration, undeniably a real manifestation of the suffering that affected many people, exact figures and precise locations are generally missing. In short, despite the available fragmented data following some scarce attempts to shed light on this issue, one largely remains in the dark with respect to the current size and distribution of the Congolese population.

In order to cope with this problem, different survey agencies have tried to correct for this knowledge gap but they did so in quite different ways, with hardly any coordination between them and without making these corrections explicit. As a matter of fact, most of the national surveys executed in the DRC were part of international initiatives, often steered from abroad, each developing its own logic and methodologies, however without any related public documentation. As a result, **many versions of demographic data have been internationally validated**, while none of them explicitly refers to the (possible) differences between its own and others' methodological approaches. The result is a plethora of internationally validated, but contradicting demographic realities. These different realities are also manifest in the varying sampling frames used underneath the sampling design of each survey.

Implications

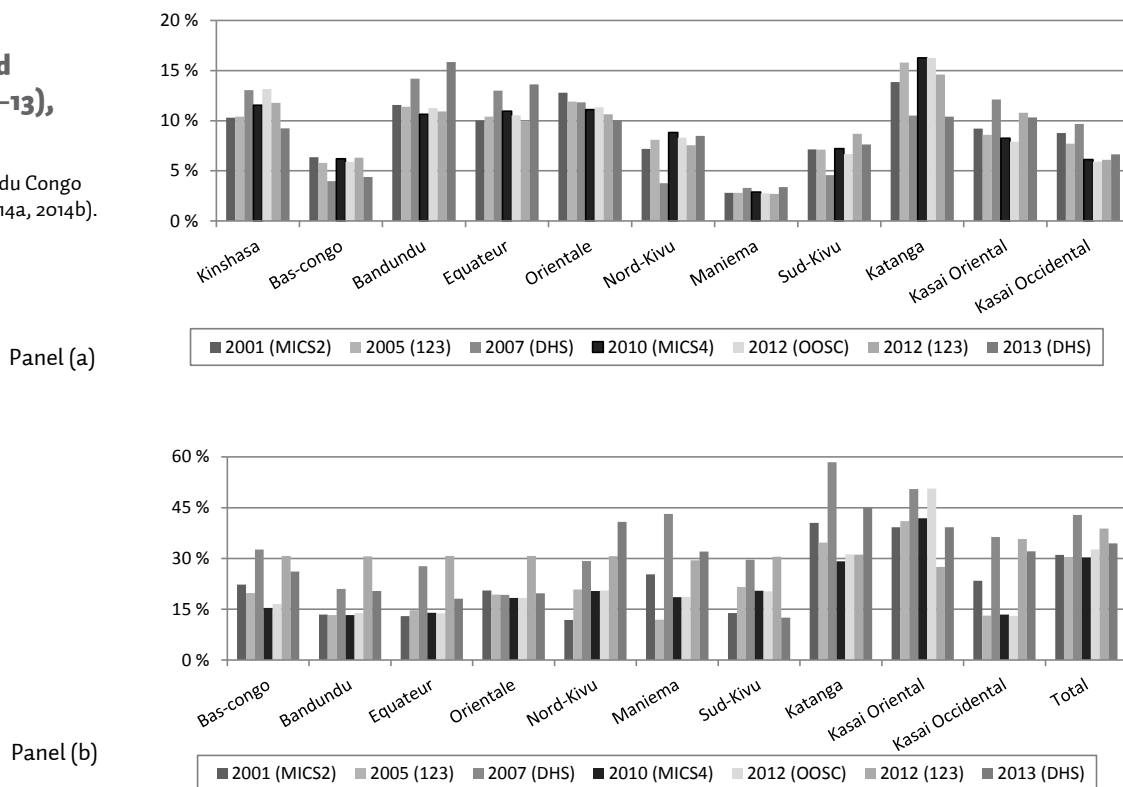
The implications of the above observations are twofold.

First of all, and most importantly, as nobody really seems to know how many Congolese today populate the DRC, **any effective or efficient policy on behalf of the government or donor community is largely excluded**. How many schools the Ministry of Education should build in the next five years in Kinshasa; how many vaccinations or food packages development partners should prepare for certain health and food-insecure regions; or how much budget the government should earmark for its current, future and retiring staff of civil servants, all depend on a knowledge of absolute figures. Yet, any presumed knowledge on this highly differs. Indeed, by using the population growth rates which implicitly informed the sampling frames of the latest seven national household surveys, the DRC would count today either 77.1 or 93.4 million people (or some figure in-between). This difference of 16.3 million equals the estimated population of the neighboring country Zambia while 2/3rd of all countries in the world have a total population size *lower* than this margin of imprecision (World Bank, 2017). It also represents 7.5 times the estimated figure of internally displaced persons in the DRC (OCHA, 2017). Apart from these implicit survey estimates, other (official) figures circulate – which further adds to the level of imprecision. To cite one example, the [Development Indicators Analysis Unit](#) (CAID), set up by the prime ministry in 2015, implicitly makes use of a figure of 101.8 million people.

Second, it is **currently impossible to field a representative survey** in the DRC in order to deepen our understanding of the country's challenges, seize their relative importance and come up with customized solutions. Evidently, without detailed knowledge on the distribution of the Congolese population,

Figure 1. Provincial demographic weight and urbanization rate (2001–13), initial sampling frame

Source: République Démocratique du Congo (2002, 2008a, 2008b, 2011, 2013, 2014a, 2014b).



one cannot accurately assign weights to each of the sampled households or individuals, which in turn affects the correct spatial aggregation of any associated survey information, be it in absolute or relative terms. To illustrate this issue, consider Figure 1, which displays the provincial demographic weight (panel a) and urbanization rate (panel b) for the latest seven national-level household surveys.

The question which pops up immediately when looking at both panels is “which of these representative surveys are truly representative?” At best one, probably none. Indeed, the ways in which various sampling experts have filled in the demographic knowledge gap is, to say the least, quite diverse. Apart from Maniema, whose share in the overall population has largely remained stable around 3 percent, all other provinces have seen their demographic weight considerably fluctuate between surveys. Whereas these fluctuations are still limited to 2–3 percent for Bas-Congo and Orientale, they become really striking for Nord-Kivu, Katanga and Bandundu with variations around 5–6 percent. Even more salient are the irregular variations in urbanization rates: for the country as a whole and for Orientale, these fluctuations did not exceed 13 percent; but in Katanga, Nord-Kivu and Maniema, one could even observe changes of around 30 percent. In these regions, three out of ten Congolese people are statistically speaking floating in and out of the urban and rural sector of their province from one survey to another.

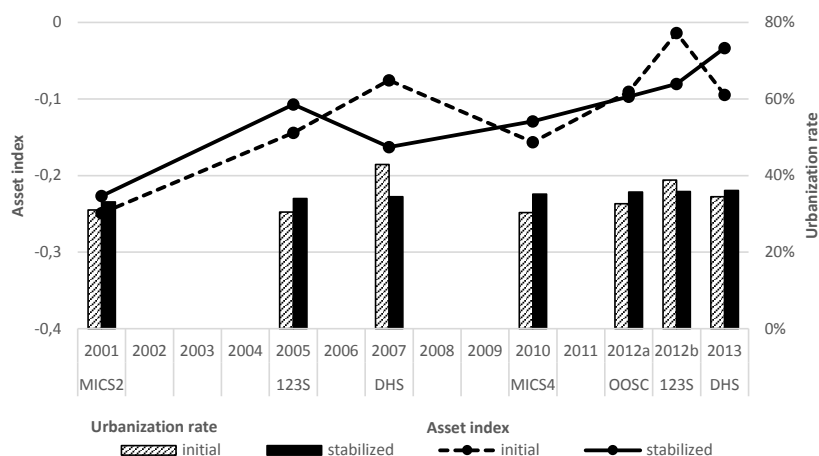
Third, and related to the former issues, **any survey-based claim about the evolution in socio-economic indicators is questionable** – given that such a claim compares the results of at least two surveys with an incoherent sampling frame. To illustrate this problem, consider Figure 2 which displays the evolution in urbanization and asset ownership between 2001 and 2013,

according to the initial and stabilized sampling frames. These stabilized sampling frames are obtained following a technique shortly described in the following section, which aims to remove the erratic nature within the initial frames by averaging out population growth and urbanization rates for each province and sector. Based on the initial weights (dashed line), asset ownership would have peaked in 2007 and again in 2012. Yet, controlling for this erratic sampling frame (solid line), the first peak probably occurred already in 2005, after which asset ownership gradually increased between 2007 and 2013, without any culmination in 2012 but with an acceleration during the last year.

Yet, by inspecting the variation in urbanization rates in detail, the two peaks in asset ownership under the initial sampling regime are probably rather driven by an overestimation of the urban population for these same years. Indeed, both in 2007 and in 2012, the urbanization rate has been estimated markedly higher than for the previous and subsequent surveys (dashed bars). In other words, without any coordination among survey agencies regarding the underlying sampling frame to be used, most trend analyses based on these surveys will be unable to disentangle real socio-economic changes from fictitious sampling fluctuations.

Figure 2. Evolution in asset ownership and urbanization rate in the DRC (2001–13)

Source: République Démocratique du Congo (2002, 2008a, 2008b, 2011, 2013, 2014a, 2014b).



What can be done?

Each year, millions of dollars are spent by the international donor community and national policymakers on household surveys to set baselines, measure progress and assess policies in developing countries. Yet, in the DRC such attempts risk to be seriously biased from their very inception, for the reasons cited above. To address this issue, only one initiative could be really conclusive: the organization of a new population census (De Saint Moulin 2009). Unfortunately, apart from multiple institutional, financial and time conditions, a census is undoubtedly even more politically sensitive than a survey, [especially in the current circumstances](#) of political instability. The politically sensitive nature directly stems from the obvious possibility that such a census would allow for a translation of the demographic weight of people, regions or ethnicities into parliamentary representation. Indirectly, it would also have an impact on certain key development indicators. For example, dividing the size of the Congolese economy in 2015 by the population estimate following the second survey of 2012 or the one conducted in 2013 (both adjusted for the year 2015) would yield a sizeable difference of 19.7%

in terms of GDP per capita. Therefore, “formally” controlling the size of the country’s population provides for an instrument to influence the assessment of any related economic policy. Besides, it also explains why political factions of different kinds stress the importance of a renewed census at various occasions, either to challenge the current power status quo or to win time – the latter given that this type of endeavor easily takes up more than a year for a country the size of the DRC.

A more feasible solution awaiting a new census is to triangulate the sampling frames of the different existing surveys with ‘real’ but limited information about the country’s demographic landscape. This is, in short, what we set out to do by making use of a combination of recent vaccination, schooling and election data, together with the national census data of 1984 (Marivoet & De Herdt, 2017). We propose a three-step procedure, which first identifies an historical and recent population reference to compute growth rates by region and sector. Using these growth rates, alternative population estimates for each survey year can be derived in a second step, which then finally allows for the computation of sampling weight adjustment factors, by confronting the initial population estimates with their alternative. Basically, this approach can be framed as a post-stratification technique. The ultimate purpose behind this harmonization process must lie in its capacity to measure socio-economic change over time, *not* to provide for yet another re-count of the Congolese population.

Besides correcting the sampling problems of surveys carried out in the recent past, our proposal can also inform sampling procedures for future surveys. Depending on the quality of other available data fragments to inform each of the three steps, the proposed technique could be further improved to allow for more spatial precision or by modifying some of the underlying assumptions adopted. Furthermore, given the generic nature of this post-stratification technique, its reach can also be easily extended to other country settings suffering from similar sampling biases.

References

Coghlan, B., Brennan, R., Ngoy, P., et al. (2008) *Mortality in the Democratic Republic of Congo: An Ongoing Crisis*. New York: International Rescue Committee.

De Saint Moulin L (2009) "Analyse du paysage sociopolitique à partir du résultat des élections de 2006. In: Trefon, T. (éd.) *Réforme au Congo (RDC): Attentes et Désillusions*. Paris: L'Harmattan, pp. 49–65.

Lambert, A. and Lohlé-Tart, L. (2008) La Surmortalité au Congo (RDC) durant les Troubles de 1998–2004: Une Estimation des Décès en Surnombre, scientifiquement Fondée à partir des Méthodes de la Démographie. Available at: www.uclouvain.be/cps/ucl/doc/demo/documents/Lambert.pdf (accessed 15 May 2016).

Little, R.J. (2004) "To Model or Not To Model? Competing Modes of Inference for Finite Population Sampling." In: *Journal of the American Statistical Association* 99(466): 546–556.

Marivoet, W. and De Herdt, T. (2017) "Tracing down real socio-economic trends from household data with erratic sampling frames. The case of the Democratic Republic of the Congo." *Journal of Asian and African Studies* (DOI: [10.1177/0021909617698842](https://doi.org/10.1177/0021909617698842)).

OCHA (2017) *Democratic Republic of the Congo: Key Figures*. Available at: <http://www.unocha.org/drc> (accessed 6 April 2017).

République Démocratique du Congo (2002) *Enquête Nationale sur la Situation des Enfants et des Femmes, MICS2/2001*. Kinshasa: Institut National de la Statistique, UNICEF, USAID.

République Démocratique du Congo (2008a) *Enquête 1-2-3 (Phase I: Emploi, Phase II: Secteur Informel et Phase III: Consommation des Ménages) 2004-5*. Kinshasa: Institut National de la Statistique.

République Démocratique du Congo (2008b) *Enquête Démographique et de Santé, République Démocratique du Congo 2007*. Calverton: Macro International, Ministère du Plan.

République Démocratique du Congo (2011) *Enquête par Grappes à Indicateurs Multiples MICS-2010, Rapport Final*. Kinshasa: Institut National de la Statistique, UNICEF.

République Démocratique du Congo (2013) *Rapport de l'Enquête Nationale sur les Enfants et Adolescents en dehors de l'Ecole (2012)*. Kinshasa: UNESCO, UKAID, UNICEF.

République Démocratique du Congo (2014a) *Enquête 1-2-3 (Phase I: Emploi, Phase II: Secteur Informel et Phase III: Consommation des Ménages) 2012-13*. Kinshasa: Institut National de la Statistique.

République Démocratique du Congo (2014b) *Deuxième Enquête Démographique et de Santé (EDS-RDC II 2013-2014)*. Rockville: Ministère du Plan et Suivi de la Mise en œuvre de la Révolution de la Modernité, Ministère de la Santé Publique, ICF International.

World Bank (2017) *World Development Indicators 2015*. Washington DC: The World Bank.