The Missing Poor*

Torsten Figueiredo Walter[§] Niclas Moneke[¶] Ana Radu[‡]

April 23, 2025

Population censuses form the basis of public resource allocation in many countries. This paper shows that census enumerators commonly face incentives to disproportionately omit members of larger households. Using microdata from 238 censuses, we estimate that this leads to undercounting in 46% of censuses. Omission is concentrated in poor countries where 0.6% of the population is missing as a result. Within affected countries, members of poor households are twice as likely to be missing as members of rich ones, leading to larger undercounts in poorer regions. We illustrate how this translates into systematic underfunding of public services in poor regions.

Keywords: Census undercounting, enumerator incentives, equity, poverty
JEL classification: C80, I32, O10

*We thank Tim Besley, Stephane Helleringer, Jeffrey Lin, Veda Narasimhan, Chris Severen, Alfred Shauri and Allison Shertzer for helpful comments. Claudia Alcarra, Zainab Amin, Frederic Cochinard, Alejandra Perez and Kristina Sisiakova provided excellent research assistance.

[§]New York University Abu Dhabi, Office A5-193, Abu Dhabi, UAE. Email: t.f.walter@nyu.edu
[¶]Department of Economics, University of Oxford, UK. Email: niclas.moneke@economics.ox.ac.uk
[‡]Department of Economics, University of Chicago. Email: aradu@uchicago.edu

1 Introduction

Population censuses represent a crucial source of information. They form the basis of public resource allocation, policy design and political representation in many countries. Therefore, failures of implementation can have grave consequences, including an inequitable allocation of resources and political misrepresentation.[1] Indeed, history is full of such failures and their ramifications.[2]

This paper studies a hitherto overlooked margin of failure in census data collection. We demonstrate that census enumerators commonly face discontinuities in effort costs of recording household members. This generates an incentive for enumerators to disproportionately omit members from larger households. We identify such discontinuities in 238 censuses across 79 countries, and document bunching in household size at the corresponding cost thresholds. We estimate that the incentive leads to population undercounting in at least 46% of censuses. Omission is concentrated in low- and lower-middle-income countries where 0.6% of the population – equivalent to 23 million people today – is missing as a result. Within affected countries, members of poor households are twice as likely to be missing as members of rich ones, as poorer households tend to be larger than richer ones. This selective undercounting leads to inequity in the allocation of public resources. Using an example from Tanzania, we illustrate how public services in poor areas are systematically underfunded as a consequence of enumerator incentives.

Across the world, census data is most commonly collected through face-to-face interviews by enumerators going from door to door.[3] For each household, enumerators are instructed to record all household members on a pre-printed roster with a fixed number of slots, one per household member. If the number of household members exceeds the number of slots, enumerators have to use a second enumeration form. This is costly because it involves at minimum the re-entering of household identifiers on the second form, and at maximum requesting a second form from a supervisor.

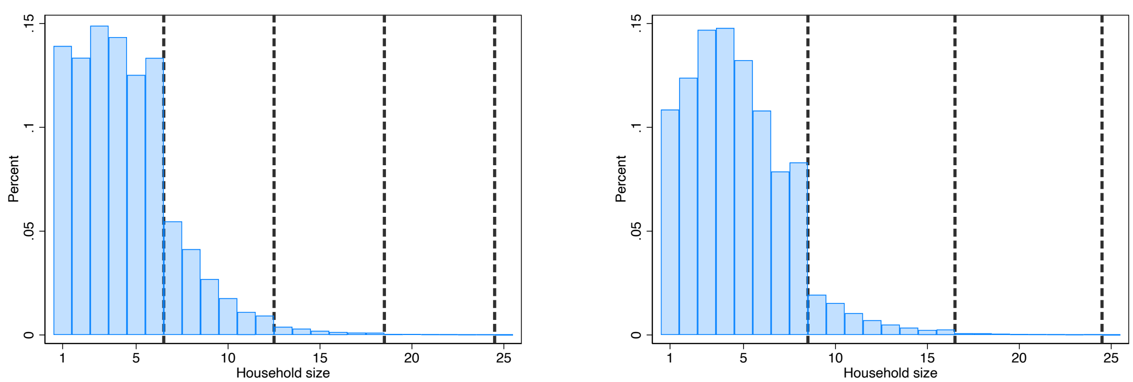
An example illustrates how this discontinuity in effort cost shapes the recording of household members: In the Tanzanian Population and Housing Census from 2002,

[1] cf. Serra and Jerven (2021) on the Nigerian census, "a politically explosive issue" due to its use as "the basis of parliamentary representation and the allocation of amenities and social services".

[2] Steckel (1991) provides an overview and history of census accuracy in the US.

[3] Two thirds of the censuses recorded in the IPUMS-International data catalogue are conducted using such 'direct enumeration' methodology. Self-enumeration and register-based censuses are only observed in some of the richer countries.

the roster included six slots for household members on the first form. Therefore, to record all members in households with seven members or more, an enumerator would require at least two forms. In contrast, the Tanzanian Population and Housing Census 2012 included eight slots on the roster, meaning enumerators needed a second form for any households with nine or more members. Figure 1 shows the resulting distribution of household sizes in either year. Dashed lines indicate the number of pre-printed slots per census form. In both census years, household sizes are bunched at full enumeration forms. Whereas bunching is prevalent at household size six in 2002, it is observed at household size eight in 2012. We interpret this as *prima facie* evidence that household sizes are systematically manipulated in response to variation in enumerators' effort cost. Enumerators are more likely to stop recording household members once a given form is full to avoid the cost of having to start a new form.



(a) Population census 2002

(b) Population census 2012

Figure 1: Household size distribution in Tanzanian Population Censuses

This figure shows the household size distributions in the 2002 and the 2012 Tanzanian Population Census. Dashed vertical lines indicate multiples of the number of pre-printed slots in the household roster (6 in 2002 and 8 in 2012).

Beyond this motivating example, we provide systematic evidence on the omission of household members from censuses due to enumerator incentives by leveraging microdata from 238 population censuses around the world, including a total of 560 million people in 138 million households. Compiled from the IPUMS-International database, these censuses were conducted across 79 countries between 1960 and 2020. We manually code the number of slots for household members per enumeration form for each census to identify discontinuities in enumerator effort cost. We document pervasive bunching in household size distributions at these discontinuities.

To identify whether and how many household members are missing from censuses affected by such bunching, we first estimate the number of excess households at the cost discontinuity under conservative assumptions about the household size distribution. Second, we set the counterfactual household size of excess households equal to the median empirically observed household size above the discontinuity. Finally, we validate our approach using placebo tests.

We estimate that as a result of enumerator incentives to stop recording once the first enumeration form is full, in 46% of censuses at least 0.1% of the population is omitted. Omission is concentrated in low- and lower-middle-income countries where 0.6% of the population is omitted. In contrast, in upper-middle- and high-income countries, only 0.1% of the population is estimated to be missing.

We leverage a small subset of nine censuses that were implemented using computer-assisted personal interviewing as a placebo. In these censuses, the marginal cost of adding members is constant throughout the household size distribution because an arbitrary number of members can be added simply by the click of a button in the digital enumeration form. Reassuringly, we do not find any evidence of bunching in the household size distribution in these censuses.

Who are the missing and how are they distributed across space within countries? We show that within affected countries, members of poor households are on average twice as likely to be missing as members of rich households. A second dimension of selection is location: we estimate that in rural areas 50% more people are missing than in urban areas. This selective undercounting also implies distortions in the relative populations of subnational administrative units. Indeed, we find that undercounting varies significantly across first-tier administrative units within countries. In the average census, omission in the subnational unit at the 90th percentile of the missingness distribution exceeds omission in the subnational unit at the 10th percentile by 0.7 percentage points.

This selective undercounting leads to inequity because population counts form the basis for the geographic allocation of public resources in many countries. In high- and low-income countries alike, central government funds are commonly distributed across subnational administrative units based on population. Additionally, the roll-out of centrally planned infrastructure is often informed by the spatial distribution of population.[4]

Since undercounting is positively correlated with poverty rates across subnational

[4]For example, see Asher and Novosad (2020) on the population-based allocation of roads in India.

regions within countries, the documented omission aggravates existing inequalities. We illustrate this at the example of Tanzania where formula-based transfers from the central government account for more than 90% of local government revenues. We estimate that the population of local government authorities in poorer North-West of the country was undercounted by up to 2.9% in the 2012 census while there was essentially no undercounting in the richer South and East. Since large shares of federal funds for public services, such as health, education and water, are distributed based on census population counts, this means that these services were underfunded in poorer areas relative to richer ones. In fact, we estimate that some local government authorities received 1.3% less funds than they were entitled to while others received up to 1.6% more. Thus, the omission of household members in larger (and poorer) households from the census by enumerators is anything but harmless. In contrast, it directly alters the distribution of funds for crucial public goods and services, with poorer locations receiving less than their fair share.

The distribution of population across space is also relevant for political representation within countries. In federal parliamentary democracies, for example, the size of the local population determines the number of seats apportioned to each state or province. Hence, differential undercounting directly leads to political misrepresentation. This can help explain why population censuses tend to be politically explosive exercises, attracting media coverage, public attention and political scrutiny.[5]

Local population counts may also affect representation indirectly through the establishment of political institutions. In India, for example, the establishment of village councils hinges on villages reaching a minimum size as recorded by the census. Having one's own council, in turn, translates into a higher supply of public goods, improving citizens' quality of life (Narasimhan & Weaver, 2024) .

Finally, selective undercounting in censuses undermines statistical representation in household surveys since their sampling frames are derived from censuses. However, national planning in many low- and middle-income countries relies heavily on these surveys, which may fail to adequately take the underrepresented poor into account.

This paper contributes to three strands of literature. First, we contribute to an old but still active literature on differential undercounting in censuses. Beginning with

[5]For example, the Nigerian 2006 census was accompanied by political violence, both the 2009 and 2019 Kenyan censuses were later partially annulled by the executive and judiciary, respectively. The Ethiopian 2017 census was engulfed in political scandal coinciding with the start of civil conflict, while mounting political pressure in India focuses on the start date of the much-delayed 2021 census.

Price (1947) seminal work on the 1940 US Census in which he uncovered a 15% undercount for blacks compared to a 3% undercount overall, differential undercounting by subgroups has been documented across a wide range of settings (Mulry & Spencer, 1993; Gumbo, 2016b; Kumar, 2024; Neidert et al., 2025). The underlying causes have remained underexplored, however. We document how a specific economic incentive drives disproportionate undercounting of members of larger households globally and quantify its impact on population counts across and within countries.

Second, this paper adds to a broad literature on the design and implementation of data collection. While the role of enumerators in the data collection process has increasingly been investigated, the literature has largely focused on how characteristics of enumerators affect the collected data (B. West & Blom, 2017; Di Maio & Fiala, 2020). Following Figueiredo Walter and Moneke (2025), we examine how enumerator incentives generated by basic data collection protocols, such as the ones used for population censuses, affect sample composition. Moreover, we confirm practitioners' concerns about enumerator manipulation of census enumeration (Price, 1947; K. K. West & Fein, 1990; Martin & de la Puente, 1993) and show that the resulting distortions matter economically.

Third, this paper relates to a rich literature on missing people across low- and middle-income countries (Sen, 1990; Sen, 1992; Qian, 2008; Foster & Rosenzweig, 1999; Ray & Anderson, 2010; Jayachandran, 2017; Anderson & Ray, 2019), insofar that it identifies a previously overlooked mechanism through which people and subgroups can go missing from official records.

The remainder of the paper is organized as follows. Section 2 describes the background and the data. Section 3 provides estimates of missing (poor) people across censuses. Section 4 illustrates implications for equity, before Section 5 concludes.

2 Background and data

2.1 Census enumeration

Census enumeration is commonly implemented through face-to-face interviews conducted at by enumerators at the doorstep.[6] For each household, enumerators complete a short enumeration form that includes a roster of all household members. Rosters

[6]While some high-income countries have introduced self-enumeration in recent decades, the traditional form of direct enumeration remains dominant across the world.

are pre-printed and as such they have a fixed number of slots for household members. If a household has more members than slots available on the form, enumerators are instructed to use a second form.[7] In this case, they need to copy all household identifiers from the first form to the second form. If enumerators do not have a spare form at hand, they have to request one from their supervisor. Figure A7 shows the enumeration form from the 2012 Tanzanian Population Census as an illustrative example, highlighting the identifiers that need to be re-entered on the second form and the checkbox that has to be ticked to indicate that a second questionnaire was used. The discrete jump in the marginal cost of effort of listing an additional household member following the completion of the pre-printed roster generates an incentive for enumerators to discontinue the enumeration of household members once the roster is full – even if the household has more members.

No pre-determined number of slots exists for household members if a census is conducted using computer-assisted personal interviewing (CAPI) instead of paper forms. Instead, a new slot is automatically created for each member. Hence, the marginal cost of adding household members is constant across household sizes.

The order in which household member are to be listed in household rosters is often explicitly specified in enumerator manuals, with little variation across time and countries. The standard listing order begins with the household head, followed by their spouse or partner, their unmarried children (often from the eldest to the youngest), their married children with their partners and their children, other relatives and finally non-relatives, including domestic employees. This provides an indication of the types of household members that are likely to be omitted if no second form is filled for large households.

Enumerator contracts vary greatly across countries. In some countries, census enumeration is conducted by teachers (e.g., Ecuador 2010, Philippines 2000) as part of their job. Sometimes teachers are supported by students in this endeavor (e.g., Costa Rica 1973, Zambia 1990). In many others, enumerators are recruited on short-term contracts (e.g., South Africa 1996, Brazil 2010). We argue that independent of the contractual details, the discrete jump in the marginal cost of effort of listing an additional household member following the completion of the pre-printed roster generates an incentive for enumerators to stop the enumeration of household members.

[7] See Appendix A.1.1 for examples of enumerator instructions on how to proceed if the number household members exceeds the number of pre-printed slots.

2.2 Census data

This paper leverages the universe of modern population census microdata available from IPUMS-International (Ruggles et al., 2024), comprising 239 censuses from 79 countries.[8] Figure A1 illustrates the geographic coverage of the data. Overall, it includes more than 138 million households globally enumerated between 1960 and 2020. The key variable for our analysis is household size which we construct by counting the number of household members recorded in each household.

We complement this data with manually coded information on the number of slots in the household roster, based on the enumeration form and the enumeration manual. The most common number of slots is 10, observed in one third of censuses. Other common thresholds are 9, 8, and 6. Finally, we record for each census whether it was paper-based or computer-assisted. Overall, nine of the 238 censuses in our data were implemented using CAPI.

3 Missing people

3.1 Empirical strategy

To identify whether and how many household members are missing from censuses, we proceed in two steps. First, we estimate the excess mass of households at the number of pre-printed slots. To this end, we assume the household size distribution continues to fall at the same rate as it is falling to the left of the threshold. In practice, this is a conservative assumption because the rate at which we see household size distributions fall to the left of the threshold is, if anything, increasing with household size – meaning that we are likely to underestimate the number of excess households at the threshold.[9]

Second, we assume that the true size of excess households is equal to the empirically observed median household size among households with more members than the number of pre-printed slots. We believe that the median provides a reasonable approximation which is not sensitive to outliers in the distribution.

We verify the validity of our approach in a subset of censuses without discontinuities in the effort cost of recording household members across household sizes. As

[8]IPUMS-International provides microdata, typically 10% samples, from a total of 354 population censuses. We disregard 115 of these for reasons detailed in Appendix A.2.1.

[9]Note that the behavior of household size distributions to the right of the threshold is not informative about the true distribution because it is distorted by the very incentive we study.

detailed in Appendix A.3, we indeed estimate omission to be very close to zero in 96% of cases.

3.2 Missing people across censuses

We find that the omission of household members varies systematically with GDP per capita across censuses. As Figure 2 shows, omission is a more serious concern in poorer countries. In low- and lower-middle-income countries, on average 0.6% of the population is missing due to the omission of members from large households. In upper-middle- and high-income countries, the missing population only amounts to 0.1% on average. The estimates for all censuses are provided in Table A1 (column 3) and plotted against GDP per capita in Figure A2.

Importantly, we do not detect any bunching in household size distribution of any of the nine censuses in our data using CAPI, as shown in Figure A4. This lends further to support to our hypothesis that the jump in the marginal cost of recording an additional household member at the end of the pre-printed roster generates an incentive for enumerators to terminate enumeration even if households have more members. It also raises hopes that the source of omission studied in this paper will be eliminated with with technological progress in the near future. At the same time, however, it is important to note that most censuses are still conducted on paper and in fact, we do not see decline, but a slight increase in missing population over time in our data (see Figure A3).

In Appendix A.4, we discuss alternative mechanisms that could explain bunching at the number of pre-printed lines. We do not find empirical support for any of them.

Our findings imply that national per capita statistics based on census population figures are slightly upward biased, especially in poorer countries.

3.3 Missing people across households

What types of households are most affected by the omission of household members? We compare the share of missing household members between poor and rich as well as urban and rural households.[10] We leverage household-level information on asset ownership to categorize households into poor and rich. Using principal component

[10]Note that census does not provide much additional household-level information. Therefore, our heterogeneity analysis is limited to these two characteristics.

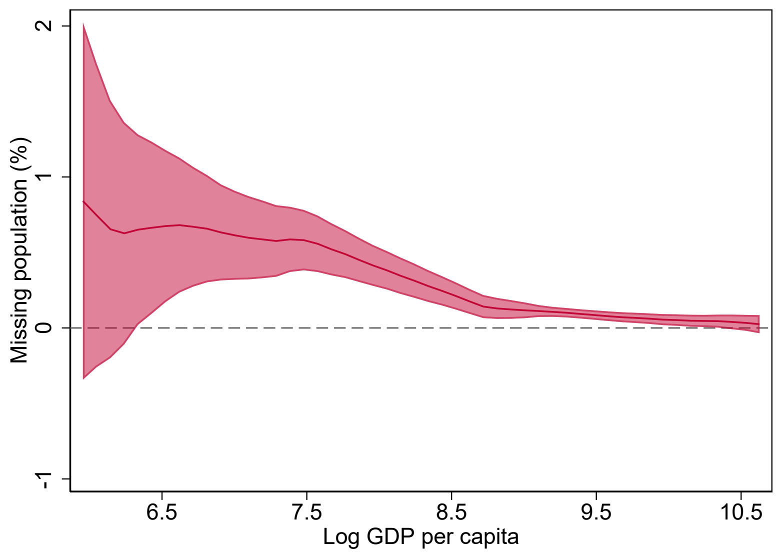


Figure 2: Missing people across censuses

This figure plots missing people against GDP per capita across censuses. The line indicates a local polynomial fit and the shaded area is the 95% confidence interval.

analysis, we build a wealth index and group households into wealth quintiles based on the resulting factor score.[11]

We find that the omission of household members is more common in poor households from the bottom wealth quintile than in rich ones from the top wealth quintile. As Figure 3 shows, in 139 out of 199 censuses for which we have adequate household asset data, the share of missing poor is weakly greater than the share of missing rich. Within the subset of censuses with a total missing population of 0.5% or more, the poor are nearly twice as likely to be missing.

Comparing the share of missing people between urban and rural households, we find that undercounting is more serious in rural areas - in line with the higher prevalence of poverty in these areas. As Figure A5 shows, in 109 out of 172 censuses for which we observe the locality of households, the share of missing rural exceeds the share of missing urban. In the subset of censuses with at least 0.5% missing people overall, members of rural households are 50% more likely to be missing than urban ones – implying an upward bias in urbanization rates in these settings.

[11]We follow the same approach used by the Demographic and Health Survey and the Multiple Indicator Cluster Survey to generate wealth quintiles. We restrict our sample to censuses that report ownership of at least four assets, corresponding to a total of 199 censuses.

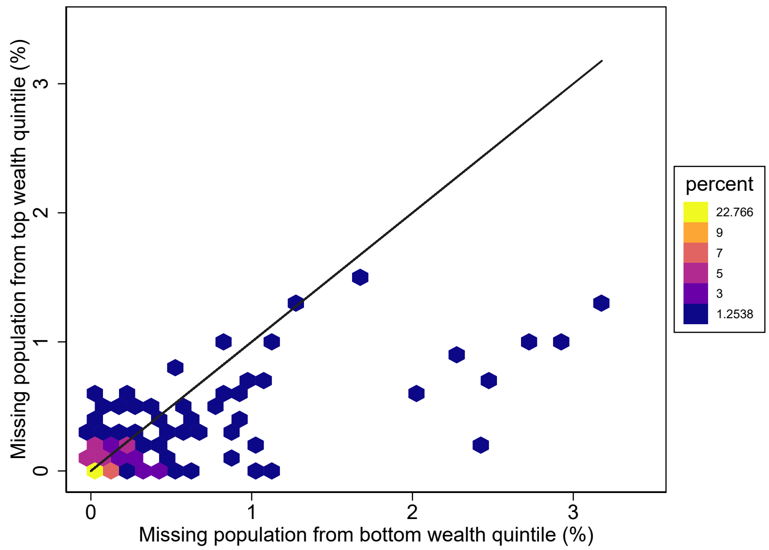


Figure 3: Missing people in poor vs rich households

This figure displays the share of missing people from rich versus poor households across population censuses. Rich households are from the top wealth quintile, poor ones from the bottom wealth quintile. Colors indicate the share of censuses falling into each hexagon. Two outliers with a large share of missing poor are omitted: Venezuela 2001 (6.2% missing poor and 2.1% missing rich) and Ethiopia 2007 (4.2% missing poor and 0% missing rich).

4 Implications for equity

4.1 Misallocation and misrepresentation

The systematic omission of household members from larger households does not only have implications for aggregate national statistics. It also has important distributional consequences because subnational population figures from the census are an important determinant of geographic allocation of public resources in many countries.

Intragovernmental transfers to local governments are often based on population. For example, in the US the geographic allocation of federal and state funds is based on census population counts (Neidert et al., 2025). In South Africa, the central government distributes grants for basic services across provinces based on population (Gumbo, 2016a). In Brazil, transfers from the national government to states and municipalities are a function of local population (Rocha, 2019).

Additionally, centrally planned infrastructure is often allocated across localities based on population. Famous examples include the Pradhan Mantri Gram Sadak Yojana program for rural road construction in India that in some states connected villages with more than 500 (or 1,000) inhabitants according to the 2001 census (Asher

& Novosad, 2020) and the Rajiv Gandhi Grameen Vidyutikaran Yojana electrification program in 27 Indian states that connected villages with more than 300 inhabitants according to the same census (Burlig & Preonas, 2024).

Selective undercounting does not only have direct implications for the allocation of public resources, it also affects political representation. In federal systems, the number of parliamentary seats apportioned to each subnational region is often based on census population counts. For example, in the US the seats in the House of Representatives are distributed across states based on population. Similarly, in Nigeria census counts of state populations form the basis of parliamentary representation (Serra & Jerven, 2021). In such systems, disproportionate undercounting of members from large households leads to their political underrepresentation.

In other contexts, the establishment of local governments itself is a function of local population. For example, in the Indian state of Uttar Pradesh, home to 241 million people, only villages with a census-enumerated population above 1,000 inhabitants are entitled to their own local government (gram panchayat). This, in turn, improves access to public goods, as shown by Narasimhan and Weaver (2024).

Finally, selective undercounting undermines the representation of large households in household surveys. These typically employ a two-stage clustered random sampling approach to generate a representative sample of the population. In the first stage, census enumeration areas are drawn with probability proportional to the population recorded in the last census. In the second stage, a given number of households is randomly drawn in each selected enumeration area. The disproportionate omission of members from larger households in the census thus leads to an underrepresentation of large households in surveys for two reasons. First, the probability of drawing enumeration areas with many large households is downward biased because their population is (more) underestimated. Second, the sampling weights for selected enumeration areas with many large households are downward biased.

Underrepresentation in household surveys matters because these surveys provide a key input to national planning – especially in poorer countries where few alternative data sources exist. The disproportionate omission of members from larger household may thus mean that their needs are not adequately reflected in policy design.

4.2 Undercounting across subnational regions

The discussion in the previous section raises two important empirical questions. First, how large are the differences in undercounting across subnational regions within countries? Second, is undercounting systematically higher in poorer regions?

We estimate the share of missing people in each first-tier administrative unit (e.g., state or province) in each census following the approach described in section 3.1. We find substantial variation in missing people across regions within countries.[12] In the average census, the subnational region at the 90th percentile in the distribution of missing people has approximately 0.7 percentage points more missing people than the subnational region at the 10th percentile.

We match our region-level estimates of missing population with regional poverty headcount ratios from the Global Multidimensional Poverty Index 2013-2024 (Alkire et al., 2024).[13] As Figure 4 shows, poverty is positively correlated with missing people across regions. Both across and within countries, a 10 percentage point increase in the poverty headcount ratio is associated with a 0.1 percentage point increase in missing population.[14]

This implies that the disproportionate omission of members from larger households is to the detriment of the poor, effectively depriving them from equal access to public services. Additionally, it disenfranchises them in certain political systems.

4.3 Public resource allocation

We illustrate the implications of spatial variation in missing population for public resource allocation at the example of Tanzania. In Tanzania, local government authorities rely heavily on transfers from the central government. In fact, these transfers account for more than 90% of their revenues (Allers & Ishemoi, 2011; Tidemand et al., 2014). Transfers are organized in the form of formula-based sectoral grants. The exact allocation formula varies across sectors. However, population is an important determinant throughout. In the health sector, 70% of the national budget is allocated

[12] See Table A2 for a summary of within-country estimates from all censuses.

[13] See Appendix A.2.2 for details.

[14] The correlation between missing people and the poverty headcount ratio across regions is almost identical across and within countries. The coefficient from a regression of the former on the latter is 0.012 across countries and 0.011 within countries, both significant at the 1% level.

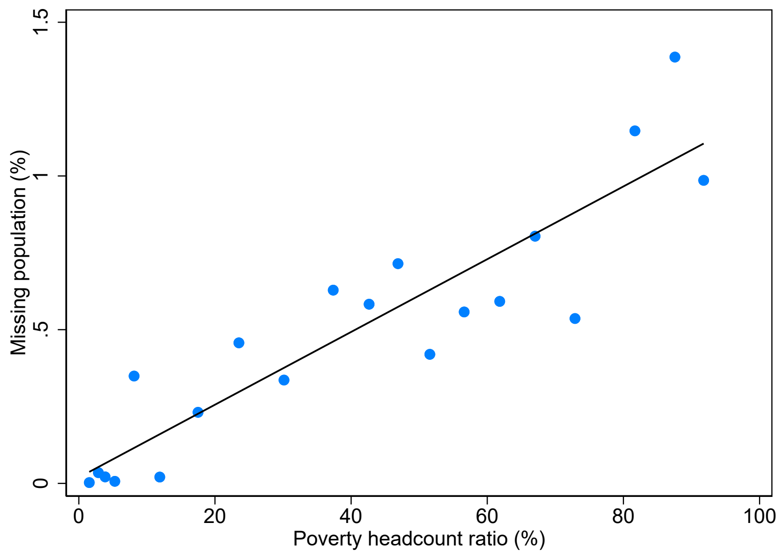


Figure 4: Missing people and poverty across subnational units

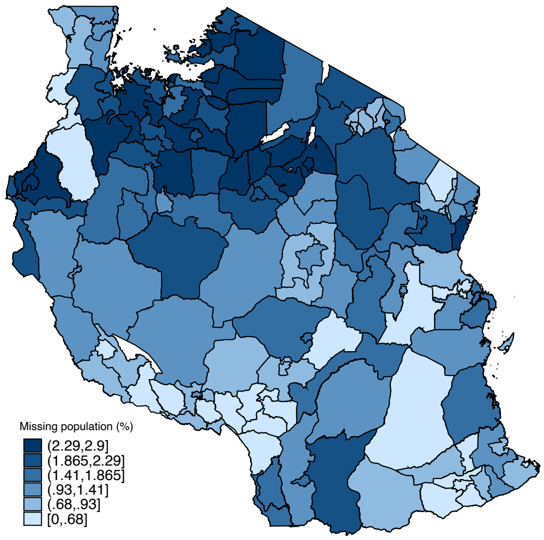
This figure shows a binscatter of estimates of missing population against poverty headcount ratios across subnational regions. Appendix A.2.2 provides details on the underlying poverty data.

across local governments based on population.[15] In the education sector, the number of school-aged children enumerated in the census is a key determinant of funding. 100% of the Block Grant for Primary Education and 70% of the Education Sector Development Grant are allocated based on this figure. Additionally, the Water Sector Development Grant and the General Purpose Grant for administration are largely distributed according to population – 70% and 50%, respectively (United Republic of Tanzania, Ministry of Finance and Economic Affairs, 2010).

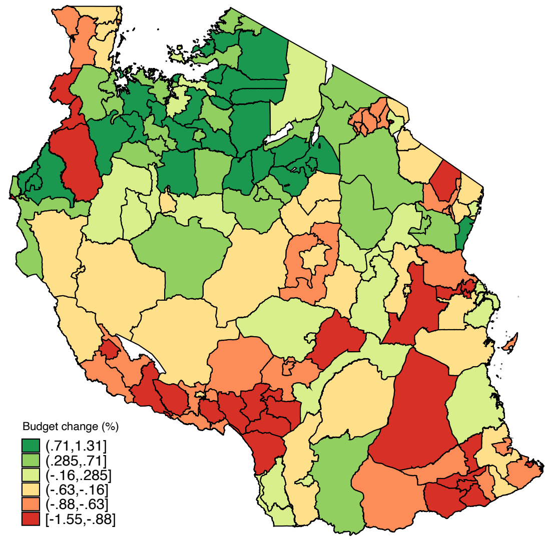
To assess the sensitivity of local government budgets to population undercounting, we estimate the missing population in the 2012 population census across local government authorities. As Figure 5a shows, there is substantial variation. At one extreme there are LGAs that do not have any missing population, at the other extreme there are LGAs with more than 2% of the population missing. Correcting population figures for the omitted household members implies significant changes in the allocation of population based funding. LGAs with few missing people lose up to 1.6% of their funding (indicated in shades of red in Figure 5b) while LGAs with a lot of missing people gain up to 1.3% (indicated in shades of green). As Figure

[15]Local government health funding is composed of a Block Grant for Primary Health, a Health Sector Basket Fund, an HIV/AIDS suvention and a Health Sector Development Grant. All of these sources of funding are allocated using the same formula.

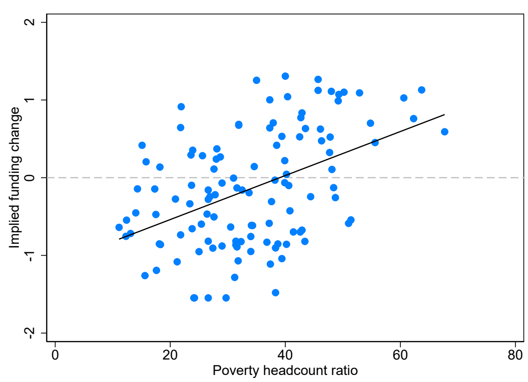
5c shows, funding gains are larger in poorer LGAs. A 10 percentage point increase in the poverty headcount ratio is associated with a 0.3 percentage point larger gain (or smaller loss) in funding.[16] This confirms that the disproportionate undercounting of members from larger households aggravates existing economic inequalities since it disproportionately affects the poor, depriving them from access to public resources.



(a) Missing population



(b) Implied funding change



(c) Implied funding change vs poverty

Figure 5: Missing population and local government funding in Tanzania

This figure displays the share of missing population across Local Government Authorities (LGAs) in Tanzania in the 2012 Population Census (Panel a) and how correcting for the missing population would affect the population-based funding of local government from intergovernmental transfers (Panel b). Finally, it plots the implied funding change against the poverty headcount ratio across LGAs (Panel c).

[16] We compile LGA-level poverty headcount ratio data from United Republic of Tanzania (2005).

5 Conclusion

Since their first documented use two millenia ago, population censuses provide essential information for the state to perform its core functions. Criticism of individual censuses, their use and abuse, abound in history, sociology, political science and demography. The economics of census data collection, however, has received considerably less attention.

In this paper, we highlight how incentives for the enumerators tasked with recording household members in population censuses affect who will be included and who will be excluded. Exploiting more than six decades of microdata from 238 population censuses across the world, we document widespread incentives for enumerators to omit members of larger households. We find that as a result, 0.6% of the population in low- and lower-middle-income countries is omitted from censuses. Within affected countries, members from poor households are twice as likely to be missing as members from rich households – because larger households tend to be poorer.

The systematic undercounting of household members from large and disproportionately poor households has important implications for equity: it aggravates existing spatial inequalities within countries because both the allocation of public resources and political representation are frequently tied to census counts of local populations.

The incentive problem at the core of this paper extends to data collection efforts other than population censuses. While we focus on discontinuities in enumerator effort costs in the listing of household members in population censuses, similar discontinuities exist in the listing of household members in household surveys. Other data collection processes that may also be affected include the listing of firms in firm censuses, where pre-printed forms are used to register all firms in a given location, and the listing of market prices in market price surveys, where pre-printed forms are used to record the prices of all goods sold in a given market.

The increasing use of tablets in data collection offers a silver lining, especially in light of the broader applicability of our findings. As demonstrated, the use of tablets eliminates the disproportionate undercounting of members from large households by removing the discontinuity in effort cost associated with moving from one paper form to another. However, changes in technology are unlikely to eliminate all incentives for enumerators to differentially exclude people from censuses. Hence, it remains crucial to understand and defuse these incentives to ensure that in the end everybody counts.

References

Alkire, S., Kanagaratnam, U., & Suppa, N. (2024). *Global multidimensional poverty index (mpi) 2013-2024*. OPHI.

Allers, M., & Ishemoi, L. (2011). Equalising spending needs of subnational governments in a developing country: The case of tanzania. *Environment and Planning C: Government and Policy*, *29*, 487–501.

Anderson, S., & Ray, D. (2019). Missing Unmarried Women. *Journal of the European Economic Association*, *17*(5), 1585–1616.

Asher, S., & Novosad, P. (2020). Rural roads and local economic development. *American Economic Review*, *110*(3), 797–823.

Burlig, F., & Preonas, L. (2024). Out of the darkness and into the light? development effects of rural electrification. *Journal of Political Economy*, *132*(9), 2937–2971.

Di Maio, M., & Fiala, N. (2020). Be wary of those who ask: A randomized experiment on the size and determinants of the enumerator effect. *World Bank Econ Rev*, *34*(3), 654–669.

Figueiredo Walter, T., & Moneke, N. (2025). Endogenous sample selection. *NYU Abu Dhabi mimeo.*

Foster, A. D., & Rosenzweig, M. R. (1999). *Missing Women, the Marriage Market and Economic Growth*, Brown University and University of Pennsylvania.

Gumbo, J. D. (2016a, April). Service delivery implications of census undercount in sub–saharan africa.

Gumbo, J. D. (2016b, March). *Undercounting controversies in south african censuses.*

Jayachandran, S. (2017). Fertility decline and missing women. *American Economic Journal: Applied Economics*, *9*(1), 118–39.

Kumar, V. (2024). Data, Development and Democracy: The Political Economy of the Indian Census. *Review of Development and Change*, *29*(2), 145–155.

Martin, E., & de la Puente, M. (1993). *Research on sources of undercoverage within households* (SM93-03). U.S. Census Bureau. Washington, DC.

Mulry, M. H., & Spencer, B. D. (1993). Accuracy of the 1990 census and undercount adjustments. *Journal of the American Statistical Association*, *88*, 1, 080–91.

Narasimhan, V., & Weaver, J. (2024). Polity size and local government performance: Evidence from india. *American Economic Review*, *114*(11), 3385–3426.

Neidert, L., Farley, R., & Morenoff, J. (2025). How census undercount became a civil rights issue and why it is increasingly important. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *11*, 26–43.

Price, D. O. (1947). A check on underenumeration in the 1940 census. *American Sociological Review*, *12*(1), 44–49.

Qian, N. (2008). Missing women and the price of tea in china: The effect of sex-specific earnings on sex imbalance. *The Quarterly Journal of Economics*, *123*(3), 1251–1285.

Ray, D., & Anderson, S. (2010). Missing women: Age and disease. *Review of Economic Studies*, *77*, 1262–1300.

Rocha, A. (2019, September). *Brazilian federal transfers to states and municipalities: A quick reference guide*. Nucleo de Estudos e Pesquisasda Consultoria Legislativa, Senado Federal.

Ruggles, S., Cleveland, L., Lovaton, R., Sarkar, S., Sobek, M., Burk, D., Ehrlich, D., Heimann, Q., & Lee, J. (2024). *Integrated public use microdata series, international: Version 7.5*. Minneapolis, MN, IPUMS.

Sen, A. (1992). Missing women. *BMJ*, *304*(6827), 587–588.

Sen, A. (1990, December). *More than 100 million women are missing*.

Serra, G., & Jerven, M. (2021). Contested numbers: Census controversies and the press in 1960s nigeria. *Journal of African History*, *62*(2), 235–253.

Steckel, R. H. (1991). The quality of census data for historical inquiry: A research agenda. *Social Science History*, *15*(4), 579–599.

Tidemand, P., Sola, N., Maziku, A., Williamson, T., Tobias, J., Long, C., & Tilley, H. (2014). Local government authority (lga) fiscal inequities and the challenges of 'disadvantaged'lgas in tanzania. *Final Report, Overseas Development Institute, UK*.

United Republic of Tanzania. (2005). *Poverty and human development report 2005* (PHDR 2005). Research and Analysis Working Group of the Government of Tanzania's Poverty Monitoring System. Dar es Salaam, Tanzania, Mkuki na Nyota Publishers.

United Republic of Tanzania, Ministry of Finance and Economic Affairs. (2010, March). *Assessment of the effectiveness of formula based budgetary allocation to local government authorities (lgas)*. Dar es Salaam, Tanzania.

West, B., & Blom, A. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, *5*, 175–211.

West, K. K., & Fein, D. J. (1990). Census undercount: An historical and contemporary sociological issue. *Sociological Inquiry*, *60*(2), 127–141.

A Online appendix

A.1 Enumerator instructions

A.1.1 More household members than pre-printed slots

If the number of household members in a household exceeds the number of pre-printed slots in the household roster, enumerators are instructed to use a second enumeration form. Below, we provide five illustrative examples of such instructions from enumerator manuals:

1. Chile 2002: "If there are more than 6 people in the household, use a second questionnaire."

2. Philippines 2010: "If there are more than eight (8) members, you will need to use additional booklet of CPH Form 2."

3. Russia 2010: "If you find out that the dwelling accommodates more than ten permanent occupants or more than four temporary occupants you will have to rearrange the questionnaires and to complete separate Forms S and P per each dwelling separately."

4. South Africa 1996: "If there are more than nine (9) people, please make use of additional questionnaires."

5. Zambia 2010: "The household listing has enough space for 8 persons. If a household has more than 8 persons, use a continuation questionnaire."

A.1.2 Ordering of household members

Enumerator manuals typically provide instructions on the order in which household members are to be listed. Below, we provide five illustrative examples:

1. Brazil 2000: "In each household, members shall be recorded in the following order: household head, spouse or partner, children and step children (in decreasing order of age), parents and parents-in-law, grandchildren, siblings, other relatives, non-relatives, domestic employees and relatives of domestic employees."

2. Fiji 1996: "Start with the head and his wife and unmarried children, beginning with the eldest and working down to the youngest. Then enter married children and their spouses and children. Then list other relatives and their wives and children who were in the household on census night. Finally list those who are not related to the head or anyone else in the household."

3. Myanmar 2014: "List members of the household by nuclear family, starting with the head and husband/wife and unmarried children, beginning with the eldest and working down to the youngest. Then list the names of each married child with spouse and their children. This must be done for every married child/couple in the household. Next, list the names of other relatives and non-relatives, in that order including visitors."

4. Uganda 1991: "Start with the head and his wife and unmarried children, beginning with the eldest and working down to the youngest. If a man has more than one wife and if all live and eat together, list each wife and her unmarried children in turn. Then enter married children and their spouses and children who spent census night with the household. Then list other relatives and their wives and children who were in the household on census night. Finally list those who are not related to the head or anyone else who spent census night with the household."

5. Ukraine 2001: "Husband and wife are recorded one after one; children (including grown up not married children) are recorded after the parents; if there are several married couples in the household firstly one married couple and their children are recorded and then another married couple and their children are recorded, etc.; household members that are not in relative or law relations with other household members are recorded ultimately."

A.2 Data

A.2.1 Selection of population censuses

IPUMS-International provides access to microdata, typically a 10% random sample, from 354 population censuses conducted between 1960 and today.[17] We use 238 of these censuses for our analysis. We disregard 116 censuses because they are not suitable for our analysis for the following reasons:

1. In 26 datasets, we cannot measure household size. This is either because persons are not organized into households or because the data does not contain all members for each sampled household.

2. 63 censuses are either register-based or self-administered.

3. For 10 censuses the documentation is insufficient to determine the number of pre-printed slots in the household roster.[18]

[17] We do not include historical censuses collected before the World Wars that are also available on IPUMS. We exclude all surveys available on IPUMS and focus only censuses.

[18] In one case, there is variation in the number of slots in the household roster across different enumeration forms.

4. For 9 censuses, the microdata does not match the listed population. This occurs in two cases: (i) The census lists both the de jure and the de facto population, but the microdata only contains one of these two populations. (ii) The census lists present usual members, absent usual members and/or visitors separately, but the microdata does not contain information on the residential status. In both cases, we do not observe the relevant population affected by the finite number of slots in the roster in the data.

5. For 8 censuses, the number of slots in the household roster is less than 6. Since our empirical strategy is not suitable for the estimation of missing population for such low thresholds, we also exclude these censuses.

A.2.2 Subnational poverty rates

We use data from the Global Multidimensional Poverty Index 2013-2024 (Alkire et al., 2024) to measure the poverty headcount ratio at the subnational region level. This poverty measure is generated from Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS). It is updated whenever a new survey is conducted in a given country. For each census in our data, we match each subnational region with the corresponding subnational region in the last preceding MPI update. This way, we obtain regional poverty information for 296 regions across 21 countries. We cannot match more regions and censuses because either there was no preceding MPI data (the first surveys underlying the MPI were conducted in 2003) or because the definitions of regions vary between censuses and the DHS/MICS.

A.3 Validation of empirical strategy

To validate our empirical strategy, we use the subset of nine censuses in our data that were conducted using CAPI. Since the marginal cost of adding household members is constant across household sizes in these censuses, there is no reason to expect any bunching. Indeed, Figure A4 demonstrates the absence of bunching in all nine household size distributions. We use our empirical strategy to estimate the missing population for hypothetical numbers of pre-printed slots in these censuses. We focus on the empirically most commonly observed thresholds 6, 7, 8, 9 and 10. As Figure A6 shows, our estimates are mostly very close to zero. Only in 4% of cases, we estimate a missing population above 0.1%.

A.4 Alternative mechanisms

An alternative explanation for the bunching of household sizes at the threshold that does not imply missing population is that all members are recorded but enumeration forms are not linked at the time of data entry (perhaps because identifiers are not entered correctly). If this were the case, one would expect to observe a significant fraction of households without household heads in censuses for which we estimate a large share of missing people (because household heads are listed first on the first form). We test this in the data. We find few households without heads. In 76% of censuses, all households have a household head. In 94% of censuses, more than 99% of households have a household head. Across censuses, there is no statistically significant relationship between the share of households without a head and our estimate of missing population.

Another alternative explanation for the observed bunching at the threshold is that the second enumeration form gets lost rather than never being filled. In this case, the population would still be undercounted, but the underlying mechanism would not be enumerator shirking. It is not clear, however, why second forms should be any more likely to be lost than first forms. If the probability of losing them is equal, on the other hand, this probability should be equal to the share of households without heads - which is very small empirically, as discussed above.

A.5 Appendix Figures

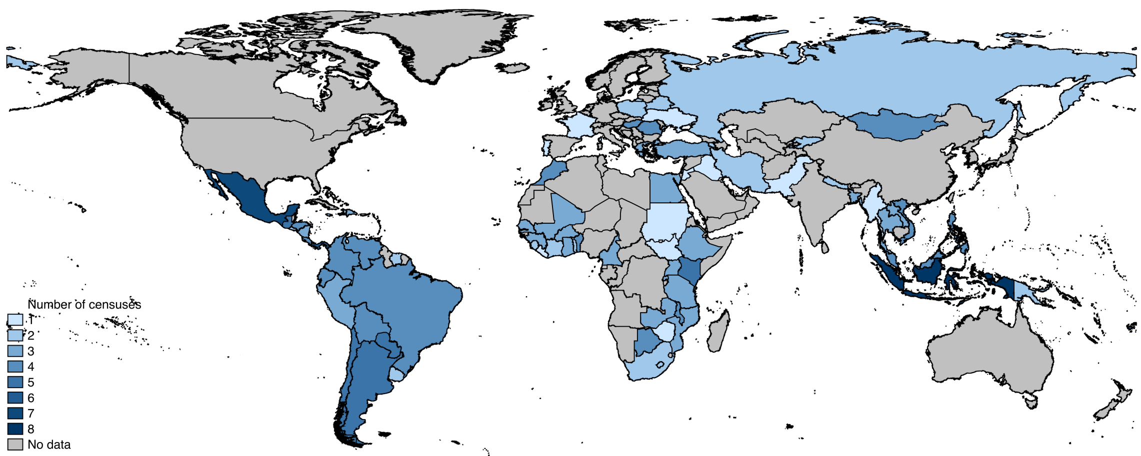


Figure A1: Geographic coverage of the data

This figure illustrates the geographic coverage of our data. It includes 238 population censuses conducted across 79 countries between 1960 and 2020.

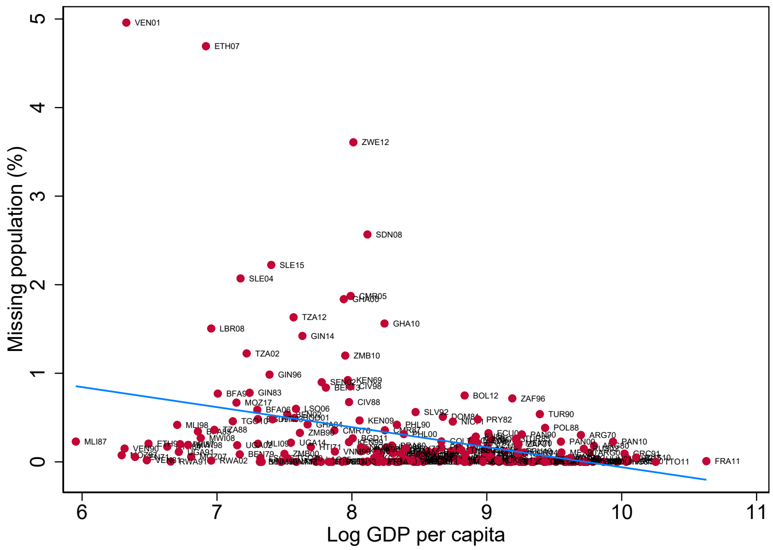


Figure A2: Missing people across censuses

This figure plots missing people against GDP per capita across censuses. Each dot represents a population census, labeled by a three-letter country code and year. The blue line indicates a linear fit.

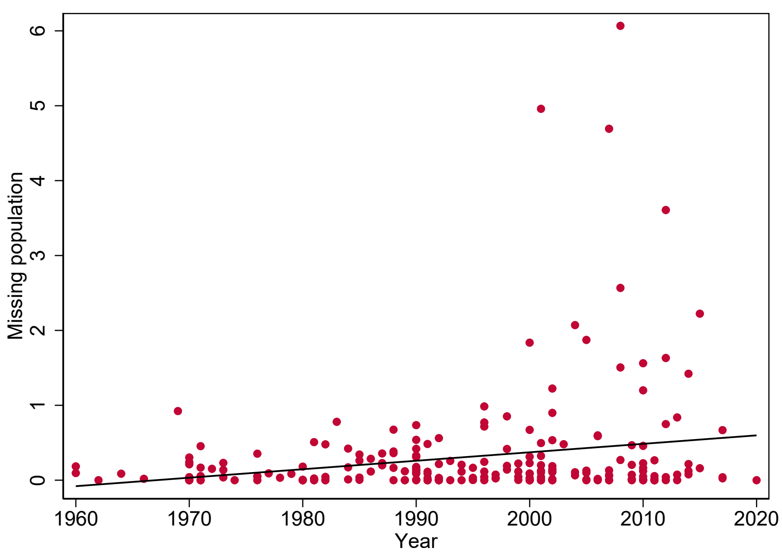
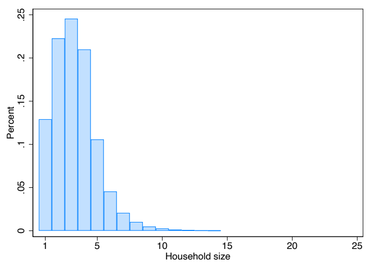
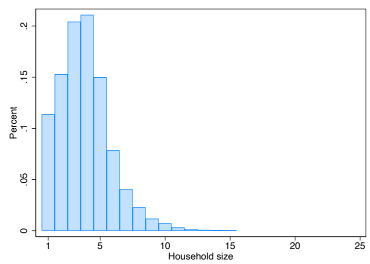


Figure A3: Missing people over time

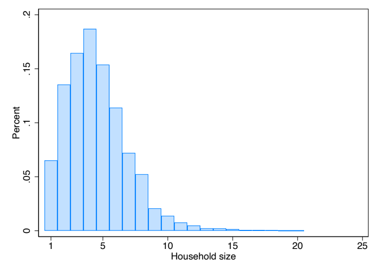
This figure displays estimates of missing population across censuses over time. The black line indicates the best linear fit.



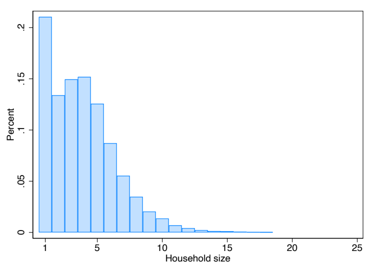
(a) Brazil 2010



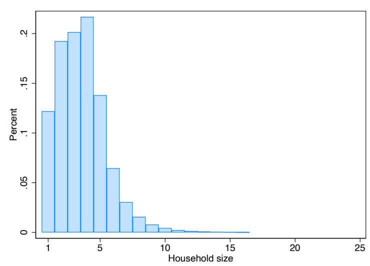
(b) Colombia 2005



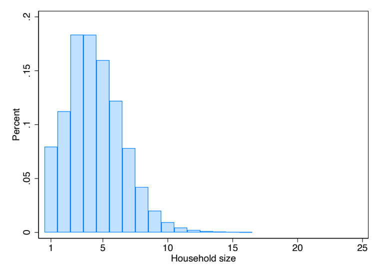
(c) Fiji 2014



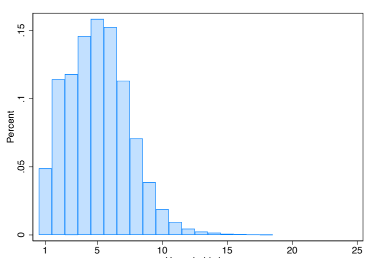
(d) Kenya 2019



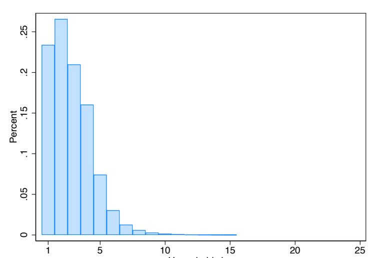
(e) Mexico 2020



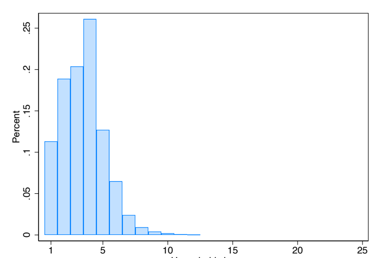
(f) Malawi 2018



(g) Palestine 2017



(h) Urguay 2011



(i) Vietnam 2019

Figure A4: Absence of bunching in censuses using tablets

This figure shows the household size distribution in all nine censuses in our data that were conducted using computer-assisted personal interviewing (CAPI).

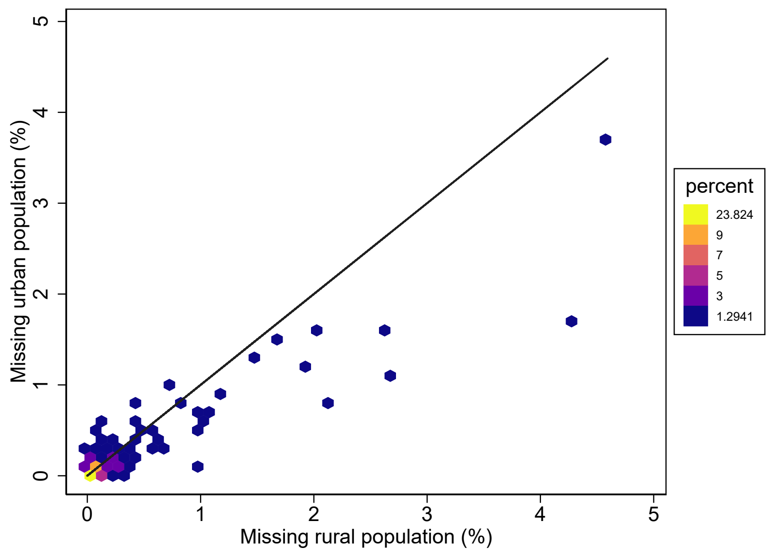


Figure A5: Missing people in urban vs rural households

This figure displays the share of missing people from rural versus urban households across population censuses. Colors indicate the share of censuses falling into each hexagon. One outlier with a large share of missing rural is omitted: Venezuela 2001 (6.5% missing rural and 4.7% missing urban).

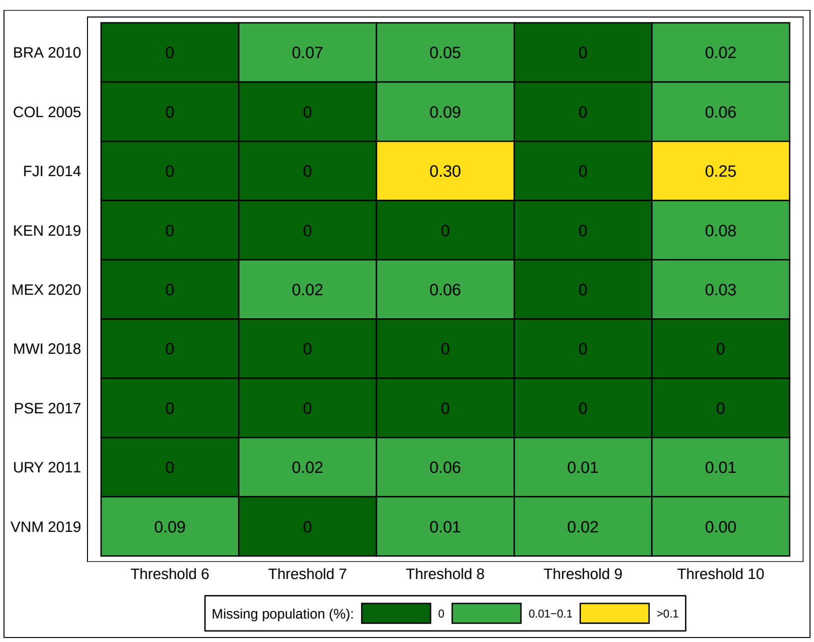
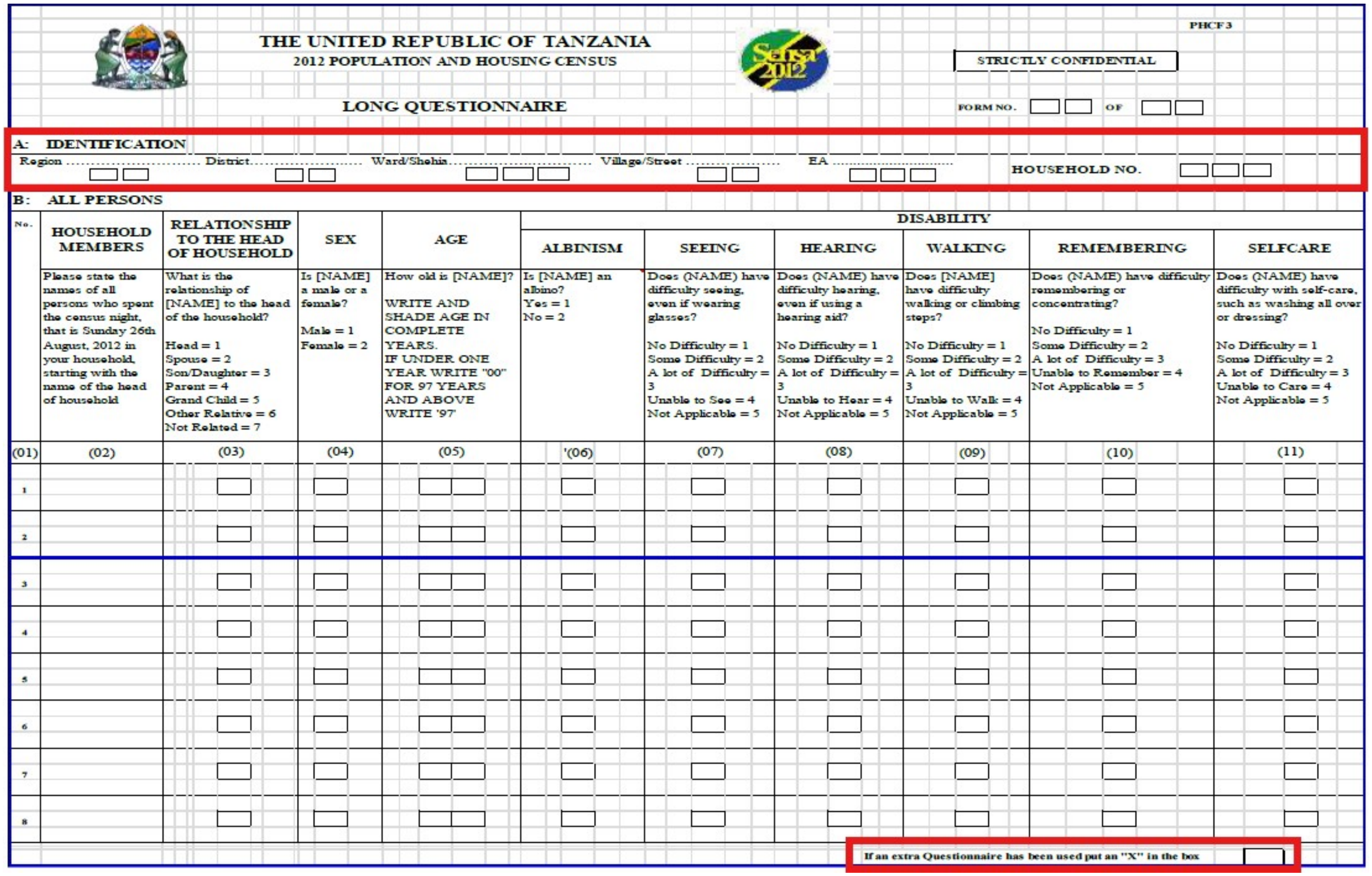


Figure A6: Placebo estimates of missing people in CAPI censuses

This figure displays estimates of missing people in CAPI censuses across different hypothetical thresholds.



PHCF 3

THE UNITED REPUBLIC OF TANZANIA
2012 POPULATION AND HOUSING CENSUS

STRICTLY CONFIDENTIAL

LONG QUESTIONNAIRE

FORM NO. ☐☐ OF ☐☐

A: IDENTIFICATION

Region .... District........................ Ward/Shehia.............................. Village/Street .... EA .... HOUSEHOLD NO.

B: ALL PERSONS

No.	HOUSEHOLD MEMBERS	RELATIONSHIP TO THE HEAD OF HOUSEHOLD	SEX	AGE	DISABILITY: ALBINISM	DISABILITY: SEEING	DISABILITY: HEARING	DISABILITY: WALKING	DISABILITY: REMEMBERING	DISABILITY: SELFCARE
	Please state the names of all persons who spent the census night, that is Sunday 26th August, 2012 in your household, starting with the name of the head of household	What is the relationship of [NAME] to the head of the household? Head = 1 Spouse = 2 Son/Daughter = 3 Parent = 4 Grand Child = 5 Other Relative = 6 Not Related = 7	Is [NAME] a male or a female? Male = 1 Female = 2	How old is [NAME]? WRITE AND SHADE AGE IN COMPLETE YEARS. IF UNDER ONE YEAR WRITE "00" FOR 97 YEARS AND ABOVE WRITE '97'	Is [NAME] an albino? Yes = 1 No = 2	Does (NAME) have difficulty seeing, even if wearing glasses? No Difficulty = 1 Some Difficulty = 2 A lot of Difficulty = 3 Unable to See = 4 Not Applicable = 5	Does (NAME) have difficulty hearing, even if using a hearing aid? No Difficulty = 1 Some Difficulty = 2 A lot of Difficulty = 3 Unable to Hear = 4 Not Applicable = 5	Does [NAME] have difficulty walking or climbing steps? No Difficulty = 1 Some Difficulty = 2 A lot of Difficulty = 3 Unable to Walk = 4 Not Applicable = 5	Does (NAME) have difficulty remembering or concentrating? No Difficulty = 1 Some Difficulty = 2 A lot of Difficulty = 3 Unable to Remember = 4 Not Applicable = 5	Does (NAME) have difficulty with self-care, such as washing all over or dressing? No Difficulty = 1 Some Difficulty = 2 A lot of Difficulty = 3 Unable to Care = 4 Not Applicable = 5
(01)	(02)	(03)	(04)	(05)	'(06)	(07)	(08)	(09)	(10)	(11)
1										
2										
3										
4										
5										
6										
7										
8										

If an extra Questionnaire has been used put an "X" in the box

Figure A7: Tanzania Population Census 2012: Enumeration form

This figure shows the enumeration form for the 2012 Tanzanian Population Census. The household identification fields and the checkbox that have to be filled if a second questionnaire is used are highlighted in red boxes.

A.6 Appendix Tables

Table A1: Missing population

Country	Year	Overall	Poor	Rich	Urban	Rural
ARG	1970	0.30				
ARG	1980	0.18	0.00	0.46	0.38	0.20
ARG	1991	0.00	0.18	0.00	0.37	0.00
ARG	2001	0.10	0.00	0.12	0.02	0.10
ARG	2010	0.01	0.00	0.37		
ARM	2001		0.00	0.00	0.00	0.00
ARM	2011	0.02	0.00	0.10	0.04	0.01
BEN	1979	0.08	0.00	0.32		
BEN	1992	0.00	0.00	0.00	0.00	0.00
BEN	2002	0.53	0.31	0.53	0.65	0.35
BEN	2013	0.84	0.60	0.94	0.85	0.81
BFA	1985	0.34				
BFA	1996	0.77	0.55	0.82		
BFA	2006	0.59	0.47	0.61	0.60	0.52
BGD	1991	0.48	0.64	0.23	0.42	0.75
BGD	2001	0.50			0.45	0.64
BGD	2011	0.27			0.25	0.31
BLR	1999	0.00	0.00	0.00	0.00	0.01
BLR	2009	0.00	0.00	0.02	0.01	0.00
BOL	1976	0.05	0.08	0.00	0.08	0.01
BOL	1992	0.03	0.05	0.03	0.04	0.03
BOL	2001	0.12	0.20	0.00	0.04	0.16
BOL	2012	0.75	0.36	0.93	1.02	0.61
BRA	1960	0.18	0.09	0.30	0.26	0.09
BRA	1970	0.04	0.04	0.07	0.05	0.02
BRA	2000	0.00	0.00	0.00	0.00	0.00
BWA	1981	0.00	0.00	0.00		
BWA	1991	0.00	0.00	0.08	0.00	0.00
BWA	2001	0.00	0.00	0.00		
BWA	2011	0.05	0.05	0.00		
CHL	1970	0.24	0.16	0.25	0.32	0.20
CHL	1982	0.00	0.00	0.05	0.00	0.00
CHL	1992	0.00	0.00	0.00	0.00	0.00
CHL	2002	0.00	0.00	0.00	0.00	0.00
CHL	2017	0.04	0.07	0.02	0.04	0.04
CIV	1988	0.67	0.80	0.48		
CIV	1998	0.85	0.52	0.74	0.95	0.74
CMR	1976	0.35	0.14	0.24		
CMR	1987	0.36	0.06	0.11	0.40	0.28
CMR	2005	1.87	0.92	2.29	2.64	1.06
COL	1973	0.23	0.05	0.35	0.42	0.11
COL	1985	0.04	0.02	0.05	0.03	0.04
COL	1993	0.26	0.15	0.48	0.43	0.19
CRI	1973	0.14	0.01	0.20	0.12	0.13
CRI	1984	0.01	0.00	0.00	0.02	0.00
CRI	2000	0.00	0.01	0.02	0.03	0.00
CRI	2011	0.00	0.00	0.02	0.02	0.00
DOM	1981	0.51	0.00	1.10		
DOM	2002	0.00	0.00	0.00	0.00	0.00
DOM	2010	0.00	0.00	0.00	0.00	0.00
ECU	1982	0.04	0.00	0.16		
ECU	1990	0.18	0.15	0.24	0.25	0.14
ECU	2001	0.32	0.21	0.35	0.14	0.36
ECU	2010	0.12	0.07	0.24	0.18	0.09
EGY	1986	0.12	0.00	0.44	0.31	0.00
EGY	1996	0.24	0.00	0.41	0.20	0.02
EGY	2006	0.00	0.00	0.00	0.00	0.00
ETH	1984	0.17	0.13	0.10	0.21	0.00
ETH	1994	0.21	0.23	0.19	0.09	0.15
ETH	2007	4.69	0.00	4.24	4.59	3.74
FJI	1966	0.02				
FJI	1976	0.05				
FJI	1986	0.29	0.33	0.43	0.11	0.55
FJI	1996	0.12	0.18	0.05		
FJI	2007	0.00	0.45	0.44	0.00	0.03
FRA	2011	0.01	0.01	0.01	0.01	0.01
GHA	1984	0.42				
GHA	2000	1.84	0.99	2.68	2.04	1.58
GHA	2010	1.56	0.74	2.46	1.92	1.21
GIN	1983	0.78	1.00	0.83	0.72	0.95
GIN	1996	0.98	0.66	1.07	1.11	0.69
GIN	2014	1.42			1.48	1.31
GRC	1971	0.00	0.00	0.00		
GRC	1981	0.00	0.00	0.00		
GRC	1991	0.09	0.00	0.00		
GTM	1964	0.09	0.29	0.00	0.07	0.17
GTM	1973	0.04			0.02	0.07

Table A1: Missing population

Country	Year	Overall	Poor	Rich	Urban	Rural
GTM	1981	0.02	0.22	0.22	0.01	0.04
GTM	1994	0.11	0.13	0.05	0.09	0.14
GTM	2002	0.12	0.08	0.11	0.12	0.12
HND	1974	0.00			0.00	0.00
HND	1988	0.16	0.25	0.13	0.09	0.24
HND	2001	0.13	0.10	0.14	0.13	0.12
HND	2013	0.07	0.00	0.10	0.05	0.10
HTI	1971	0.17	0.35	0.00	0.08	0.51
HTI	2003	0.48	0.46	0.37	0.48	0.48
HUN	1970	0.00	0.00	0.00	0.00	0.00
HUN	1980	0.00	0.00	0.00		
HUN	1990	0.01	0.00	0.01	0.00	0.02
HUN	2001	0.03	0.01	0.05		
IDN	1971	0.00	0.00	0.00	0.00	0.00
IDN	1976	0.00	0.00	0.00	0.00	0.02
IDN	1980	0.01	0.01	0.00	0.01	0.01
IDN	1985	0.00	0.02	0.01	0.00	0.01
IDN	1990	0.00	0.01	0.00	0.01	0.00
IDN	1995	0.00	0.01	0.00	0.00	0.00
IDN	2000	0.09			0.09	0.09
IDN	2005	0.01	0.00	0.00	0.00	0.00
IRN	2006	0.02	0.02	0.01	0.03	0.01
IRN	2011	0.00			0.00	0.00
IRQ	1997	0.02	0.09	0.03	0.04	0.00
JAM	1991	0.03	0.04	0.10		
JAM	2001	0.04	0.09	0.09	0.10	0.00
JOR	2004	0.11	0.01	0.22	0.24	0.06
KEN	1969	0.92				
KEN	1989	0.12	0.21	0.07	0.11	0.18
KEN	1999	0.22	0.15	0.20	0.23	0.21
KEN	2009	0.47	0.14	0.91	0.56	0.26
KGZ	1999	0.05	0.04	0.04	0.04	0.06
KGZ	2009	0.07			0.06	0.09
LAO	1995	0.02	0.09	0.03	0.04	0.00
LAO	2005	0.13	0.08	0.11	0.06	0.30
LAO	2015	0.16	0.06	0.14	0.19	0.09
LBR	2008	1.51	0.23	2.46	2.12	0.81
LSO	1996	0.04	0.00	0.09	0.09	0.00
LSO	2006	0.60	0.31	0.84	0.67	0.27
MAR	1982	0.00	0.00	0.00		
MAR	1994	0.00	0.09	0.00		
MAR	2004	0.07	0.00	0.07		
MAR	2014	0.08	0.11	0.00	0.00	0.08
MEX	1970	0.21	0.22	0.22	0.17	0.25
MEX	1990	0.11	0.20	0.13	0.10	0.04
MEX	1995	0.03	0.00	0.00	0.00	0.04
MEX	2000	0.00	0.00	0.00	0.00	0.00
MEX	2005	0.00	0.00	0.00	0.00	0.00
MEX	2010	0.00	0.00	0.00	0.00	0.00
MLI	1987	0.23	0.19	0.23		
MLI	1998	0.42	0.52	0.27	0.41	0.42
MLI	2009	0.20	0.47	0.16	0.16	0.36
MMR	2014	0.12	0.25	0.01	0.07	0.21
MNG	1989	0.00				
MNG	2000	0.00	0.00	0.00		
MNG	2010	0.00	0.00	0.00		
MNG	2020	0.00	0.00	0.00		
MOZ	1997	0.07	0.00	0.06	0.08	0.03
MOZ	2007	0.06	0.11	0.02	0.04	0.08
MOZ	2017	0.67	0.00	1.06	0.94	0.09
MUS	1990	0.09			0.07	0.12
MUS	2000	0.00			0.00	0.04
MWI	1987	0.20	0.29	0.19	0.19	0.26
MWI	1998	0.19	0.10	0.14	0.21	0.07
MWI	2008	0.27	0.47	0.09	0.24	0.42
MYS	1970	0.00	0.00	0.00	0.00	0.00
MYS	1980	0.00	0.00	0.00	0.00	0.00
MYS	1991	0.01	0.00	0.00	0.00	0.00
MYS	2000	0.00	0.00	0.00	0.00	0.00
NIC	1971	0.45	0.00	1.05		
NIC	1995	0.17	0.16	0.05		
NIC	2005	0.10	0.06	0.32	0.10	0.10
NPL	2001	0.00	0.00	0.00	0.00	0.00
NPL	2011	0.00	0.00	0.00	0.00	0.00
PAK	1998	0.14	0.34	0.00	0.09	0.24
PAN	1960	0.10	0.00	0.01	0.00	0.28
PAN	1970	0.00			0.00	0.01
PAN	1980	0.00	0.00	0.50	0.17	0.00
PAN	1990	0.31	0.00	0.59		
PAN	2000	0.23	0.23	0.35	0.22	0.20
PAN	2010	0.23	0.27	0.27	0.19	0.21
PER	1993	0.00	0.00	0.39	0.27	0.00

Table A1: Missing population

Country	Year	Overall	Poor	Rich	Urban	Rural
PER	2007	0.07	0.06	0.07	0.07	0.07
PER	2017	0.03	0.02	0.04	0.02	0.02
PHL	1990	0.42	0.45	0.37	0.42	0.41
PHL	1995	0.00				
PHL	2000	0.31	0.31	0.27		
PHL	2010	0.16	0.23	0.14		
PNG	1990	0.74	0.32	0.81	0.81	0.32
PNG	2000	0.67			0.72	0.33
POL	1978	0.03	0.00	0.00	0.00	0.00
POL	1988	0.38	0.00	0.00		
PRT	1991	0.00	0.00	0.02	0.00	0.00
PRY	1962	0.00			0.00	0.04
PRY	1972	0.15	0.00	0.21	0.28	0.00
PRY	1982	0.48	0.44	0.63	0.45	0.45
PRY	1992	0.22	0.04	0.10	0.32	0.09
PRY	2002	0.11	0.04	0.12	0.09	0.11
PSE	1997	0.03	0.10	0.00	0.04	0.03
PSE	2007	0.00	0.00	0.00	0.00	0.00
ROU	1977	0.09	0.14	0.00	0.00	0.14
ROU	1992	0.00	0.00	0.00	0.00	0.00
ROU	2002	0.01	0.00	0.01	0.00	0.00
ROU	2011		0.00	0.00	0.00	0.00
RUS	2002	0.15			0.16	0.14
RUS	2010	0.05			0.06	0.05
RWA	1991	0.00	0.11	0.00		
RWA	2002	0.01	0.16	0.01	0.01	0.01
RWA	2012	0.00	0.01	0.00	0.00	0.02
SDN	2008	2.57	1.24	3.18	4.29	1.70
SEN	1988	0.00	0.00	0.00		
SEN	2002	0.90	0.68	1.00	0.98	0.76
SEN	2013	0.00	0.00	0.00	0.04	0.00
SLE	2004	2.07	0.57	1.99	2.64	1.11
SLE	2015	2.22	1.01	2.93	2.67	1.60
SLV	1992	0.56	0.29	0.64		
SLV	2007	0.13	0.08	0.11	0.12	0.10
SSD	2008	6.07	4.91	6.09	6.15	5.49
SUR	2012	0.04	0.06	0.00		
TGO	1970	0.04				
TGO	2010	0.46	0.27	0.47	0.53	0.34
THA	1970	0.00	0.00	0.00	0.00	0.07
THA	1980	0.00	0.04	0.00	0.00	0.00
THA	1990	0.13	0.21	0.13	0.11	0.20
TTO	1980	0.00	0.35	0.00		
TTO	1990	0.00	0.00	0.14		
TTO	2000	0.00	0.00	0.00		
TTO	2011	0.00	0.00	0.14		
TUR	1985	0.26				
TUR	1990	0.54				
TUR	2000	0.03	0.00	0.04		
TZA	1988	0.36	0.23	0.40		
TZA	2002	1.22	1.30	1.27	1.19	0.90
TZA	2012	1.63	1.44	1.67	1.69	1.48
UGA	1991	0.11	0.22	0.07	0.13	0.09
UGA	2002	0.19	0.17	0.36	0.19	0.19
UGA	2014	0.22	0.20	0.27	0.21	0.22
UKR	2001	0.02			0.02	0.04
URY	1996	0.00	0.00	0.00		
VEN	1971	0.06	0.09	0.06		
VEN	1981	0.02	0.01	0.00	0.00	0.02
VEN	1990	0.15	0.12	0.17	0.31	0.13
VEN	2001	4.96	2.10	6.17	6.53	4.74
VNM	1989	0.00	0.07	0.00	0.00	0.08
VNM	1999	0.12	0.34	0.08	0.03	0.22
VNM	2009	0.01	0.62	0.02	0.00	0.22
ZAF	1996	0.72	0.20	1.04	0.98	0.50
ZAF	2001	0.20	0.09	0.26	0.26	0.15
ZMB	1990	0.33	0.14	0.20	0.40	0.21
ZMB	2000	0.09	0.06	0.07	0.11	0.05
ZMB	2010	1.20	0.99	1.09		
ZWE	2012	3.61	0.17	0.27	27.35	20.43

Table A2: Missing population by region

Country	Year	10th pctl	Median	90th pctl
ARG	1970	0.00	0.28	0.91
ARG	1980	0.00	0.19	0.54

Table A2: Missing population by region

Country	Year	10th pctl	Median	90th pctl
ARG	1991	0.00	0.00	0.17
ARG	2001	0.00	0.04	0.17
ARG	2010	0.00	0.00	0.14
ARM	2001	0.00	0.00	0.00
ARM	2011	0.00	0.00	0.10
BEN	1979	0.00	0.12	0.51
BEN	1992	0.00	0.00	0.00
BEN	2002	0.00	0.43	1.22
BEN	2013	0.42	0.80	1.26
BFA	1985	0.07	0.36	0.54
BFA	1996	0.50	0.81	1.10
BFA	2006	0.21	0.57	0.85
BGD	1991	0.37	0.47	0.67
BGD	2001	0.32	0.42	0.91
BGD	2011	0.06	0.12	0.30
BLR	1999	0.00	0.01	0.01
BLR	2009	0.00	0.00	0.00
BOL	1976	0.00	0.00	2.04
BOL	1992	0.00	0.01	0.10
BOL	2001	0.00	0.11	0.80
BOL	2012	0.36	0.90	1.85
BRA	1960	0.06	0.21	0.53
BRA	1970	0.00	0.04	0.19
BRA	2000	0.00	0.00	0.00
BRA	2010	0.00	0.00	0.00
BWA	1981	0.00	0.15	2.66
BWA	1991	0.00	0.00	1.03
BWA	2001	0.00	0.00	0.03
BWA	2011	0.00	0.05	0.46
CHL	1970	0.00	0.24	0.88
CHL	1982	0.00	0.00	0.00
CHL	1992	0.00	0.00	0.00
CHL	2002	0.00	0.00	0.00
CHL	2017	0.00	0.04	0.24
CIV	1988	0.00	0.63	1.35
CIV	1998	0.02	0.88	1.79
CMR	1976	0.00	0.28	1.73
CMR	1987	0.00	0.26	1.18
CMR	2005	0.60	1.75	3.49
COL	1973	0.00	0.27	0.82
COL	1985	0.00	0.07	0.19
COL	1993	0.11	0.27	0.64
COL	2005	0.00	0.00	0.00
CRI	1973	0.00	0.05	0.42
CRI	1984	0.00	0.01	0.21
CRI	2000	0.00	0.02	0.06
CRI	2011	0.00	0.00	0.02
DOM	1981	0.08	0.71	1.61
DOM	2002	0.00	0.00	0.00
DOM	2010	0.00	0.00	0.00
ECU	1982	0.00	0.04	0.55
ECU	1990	0.00	0.16	0.41
ECU	2001	0.00	0.23	0.68
ECU	2010	0.02	0.12	0.45
EGY	1986	0.00	0.10	0.82
EGY	1996	0.00	0.27	0.98
EGY	2006	0.00	0.00	0.04
ETH	1984	0.00	0.00	0.47
ETH	1994	0.00	0.09	1.25
ETH	2007	1.69	4.43	15.75
FJI	1966	0.00	0.09	1.12
FJI	1976	0.00	0.09	0.56
FJI	1986	0.00	0.46	0.67
FJI	1996	0.00	0.06	0.20
FJI	2007	0.00	0.00	0.09
FJI	2014	0.00	0.00	0.00
FRA	2011	0.00	0.01	0.01
GHA	1984	0.21	0.37	0.79
GHA	2000	1.34	1.53	3.87
GHA	2010	0.59	1.01	2.72
GIN	1983	0.19	0.50	1.30
GIN	1996	0.01	1.10	1.76
GIN	2014	0.68	1.49	2.19
GRC	1971	0.00	0.00	0.08
GRC	1981	0.00	0.00	0.25
GRC	1991	0.00	0.04	0.41
GTM	1964	0.00	0.06	0.20
GTM	1973	0.00	0.02	0.60
GTM	1981	0.00	0.00	0.41
GTM	1994	0.00	0.09	0.25
GTM	2002	0.00	0.13	0.31
HND	1974	0.00	0.01	0.37

Table A2: Missing population by region

Country	Year	10th pctl	Median	90th pctl
HND	1988	0.00	0.22	0.54
HND	2001	0.00	0.16	0.49
HND	2013	0.00	0.11	0.38
HTI	1971	0.04	0.19	0.23
HTI	2003	0.11	0.42	0.98
IDN	1971	0.00	0.00	0.06
IDN	1976	0.00	0.00	0.06
IDN	1980	0.00	0.00	0.04
IDN	1985	0.00	0.00	0.04
IDN	1990	0.00	0.00	0.03
IDN	1995	0.00	0.00	0.01
IDN	2000	0.01	0.06	0.18
IDN	2005	0.00	0.01	0.06
IRN	2006	0.00	0.00	0.05
IRN	2011	0.00	0.00	0.00
IRQ	1997	0.00	0.00	0.30
JAM	1991	0.00	0.06	0.37
JAM	2001	0.00	0.08	0.16
JOR	2004	0.00	0.09	0.44
KEN	1969	0.23	0.89	2.35
KEN	1989	0.03	0.09	0.51
KEN	1999	0.12	0.17	0.99
KEN	2009	0.05	0.24	3.04
KEN	2019	0.00	0.00	0.00
KGZ	1999	0.00	0.03	0.10
KGZ	2009	0.00	0.05	0.16
LAO	1995	0.00	0.00	0.30
LAO	2005	0.00	0.02	0.35
LAO	2015	0.00	0.16	0.35
LBR	2008	0.48	1.25	5.83
LSO	1996	0.00	0.05	0.30
LSO	2006	0.18	0.65	1.53
MAR	1982	0.00	0.00	2.86
MAR	1994	0.00	0.04	0.16
MAR	2004	0.00	0.05	0.57
MAR	2014	0.00	0.07	0.18
MEX	1970	0.00	0.19	0.49
MEX	1990	0.00	0.04	0.32
MEX	1995	0.00	0.01	0.12
MEX	2000	0.00	0.00	0.00
MEX	2005	0.00	0.00	0.00
MEX	2010	0.00	0.00	0.00
MEX	2020	0.00	0.00	0.00
MLI	1987	0.00	0.18	0.48
MLI	1998	0.00	0.39	0.58
MLI	2009	0.00	0.15	0.41
MMR	2014	0.00	0.09	0.26
MNG	1989	0.00	0.00	0.00
MNG	2000	0.00	0.00	0.30
MNG	2010	0.00	0.00	0.00
MNG	2020	0.00	0.00	0.00
MOZ	1997	0.00	0.02	0.18
MOZ	2007	0.00	0.05	0.15
MOZ	2017	0.00	0.41	1.35
MUS	1990	0.00	0.11	0.22
MUS	2000	0.00	0.00	0.22
MWI	1987	0.00	0.20	0.46
MWI	1998	0.00	0.15	0.37
MWI	2008	0.00	0.23	0.86
MWI	2018	0.00	0.00	0.00
MYS	1970	0.00	0.00	0.00
MYS	1980	0.00	0.00	0.01
MYS	1991	0.00	0.00	0.01
MYS	2000	0.00	0.00	0.35
NIC	1971	0.00	0.60	1.42
NIC	1995	0.00	0.16	0.41
NIC	2005	0.00	0.14	0.47
NPL	2001	0.00	0.00	0.00
NPL	2011	0.00	0.00	0.00
PAK	1998	0.00	0.14	0.51
PER	1993	0.00	0.00	0.37
PER	2007	0.00	0.05	0.17
PER	2017	0.00	0.02	0.06
PHL	1990	0.00	0.36	0.81
PHL	1995	0.00	0.00	0.22
PHL	2000	0.00	0.22	0.73
PHL	2010	0.00	0.14	0.40
PNG	1990	0.00	0.64	1.95
PNG	2000	0.00	0.25	2.07
POL	1978	0.00	0.00	0.08
POL	1988	0.00	0.00	0.05
PRT	1991	0.00	0.00	0.08

Table A2: Missing population by region

Country	Year	10th pctl	Median	90th pctl
PRY	1962	0.00	0.00	0.76
PRY	1972	0.00	0.18	0.81
PRY	1982	0.00	0.50	0.79
PRY	1992	0.00	0.14	0.56
PRY	2002	0.00	0.11	0.46
PSE	1997	0.00	0.04	0.15
PSE	2007	0.00	0.00	0.04
PSE	2017	0.00	0.00	0.00
ROU	1977	0.00	0.07	0.19
ROU	1992	0.00	0.00	0.02
ROU	2002	0.00	0.00	0.02
ROU	2011	0.00	0.00	0.00
RUS	2002	0.00	0.02	0.20
RUS	2010	0.00	0.02	0.08
RWA	1991	0.00	0.02	0.06
RWA	2002	0.00	0.02	0.12
RWA	2012	0.00	0.01	0.01
SDN	2008	1.10	1.68	5.02
SEN	1988	0.00	0.00	0.00
SEN	2002	0.64	0.89	1.31
SEN	2013	0.00	0.03	0.26
SLE	2004	0.79	1.91	3.81
SLE	2015	1.22	2.03	5.51
SLV	1992	0.06	0.67	0.94
SLV	2007	0.00	0.09	0.30
SSD	2008	1.60	4.14	11.53
SUR	2012	0.00	0.00	0.10
TGO	1970	0.00	0.14	0.20
TGO	2010	0.29	0.36	0.96
THA	1970	0.00	0.00	0.19
THA	1980	0.00	0.00	0.47
THA	1990	0.00	0.11	0.30
TTO	1980	0.00	0.00	0.59
TTO	1990	0.00	0.00	0.22
TTO	2000	0.00	0.00	0.04
TTO	2011	0.00	0.00	0.33
TUR	1985	0.00	0.19	0.61
TUR	1990	0.00	0.41	1.26
TUR	2000	0.00	0.00	0.12
TZA	1988	0.00	0.25	0.68
TZA	2002	0.29	1.13	1.58
TZA	2012	0.35	1.26	2.46
UGA	1991	0.00	0.11	0.35
UGA	2002	0.00	0.10	0.39
UGA	2014	0.00	0.15	0.58
UKR	2001	0.00	0.01	0.03
URY	1996	0.00	0.00	0.65
URY	2011	0.00	0.00	0.00
VEN	1971	0.00	0.07	0.20
VEN	1981	0.00	0.01	0.04
VEN	1990	0.00	0.16	0.43
VEN	2001	1.15	5.09	6.08
VNM	1989	0.00	0.00	0.14
VNM	1999	0.00	0.02	0.35
VNM	2009	0.00	0.03	0.29
VNM	2019	0.00	0.00	0.00
ZAF	1996	0.34	0.61	1.11
ZAF	2001	0.07	0.15	0.42
ZMB	1990	0.04	0.34	0.69
ZMB	2000	0.00	0.07	0.25
ZMB	2010	0.85	1.25	1.64
ZWE	2012	1.85	2.27	3.18