DataFirst

Measuring rural households and electricity access: A comparison of national census data and small-area health and demographic surveillance system (HDSS) data

Takwanisa Machemedze, Mercy Shoko, Mark Collinson and Martin Wittenberg

Technical Paper Series
Number 42

About the Author(s)

Takwanisa Machemedze: DataFirst, University of Cape Town
Mercy Shoka: School of Public Health, University of the Witwatersrand; Demographic & Population Statistics, and Health & Vital Statistics, Statistics South Africa
Mark Collinson: Population and Public Health, School of Public Health, University of the Witwatersrand; MRC/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt); South African Population Research Infrastructure Network (SAPRIN), Department of Science and Technology and South African Medical Research Council
Martin Wittenberg: School of Economics, and DataFirst, University of Cape Town

Recommended citation

Machemedze, T., Shoko, M., Collinson, M.A., Wittenberg, M. (2022). Measuring rural households and electricity access: A comparison of national census data and small-area health and demographic surveillance system (HDSS) data. A DataFirst Technical Paper 42. Cape Town: DataFirst, University of Cape Town.

DataFirst, University of Cape Town, Private Bag, Rondebosch, 7701, Tel: (021) 650 5708, Email: info@data1st.org/support@data1st.org

Measuring rural households and electricity access: A comparison of national census data and small-area health and demographic surveillance system (HDSS) data

Machemedze, T.[1], Shoko, M.[2], Collinson, M.A.[3], Wittenberg M.[4]

DataFirst Technical Paper 42
University of Cape Town
November 2022

Abstract

Progress on achieving the Sustainable Development Goals requires high quality measurement. Too few attempts are made to assess the accuracy of existing measurements and how it changes over time. We compare household counts and electrification rates for the Agincourt Health and Demographic Surveillance Site (HDSS), as measured in the 1996, 2001 and 2011 national censuses and in the database of the HDSS. The household measurements in the two systems agree within a few percentage points in 2001 and 2011 but show much bigger divergences in 1996. The population counts also show impressive agreement, with perhaps some over-enumeration of older males in the national census. Overall, survey and census information seem to provide accurate measures of population access to electricity.

[1] DataFirst, University of Cape Town
[2] School of Public Health, University of the Witwatersrand; Demographic & Population Statistics, and Health & Vital Statistics, Statistics South Africa
[3] Population and Public Health, School of Public Health, University of the Witwatersrand; MRC/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt); South African Population Research Infrastructure Network (SAPRIN), Department of Science and Technology and South African Medical Research Council
[4] School of Economics, and DataFirst, University of Cape Town

1. Introduction

What progress has been made in eradicating poverty? How many people are still relying on unsafe fuels for cooking? The sustainable development goals (SDGs) are central to the discussions about policy challenges facing the world. A key issue is accurate and appropriate measurement. In this context the UN Secretary General's Independent Expert Advisory Group (IEAG) called for a "Data Revolution for Sustainable Development" (IEAG 2014). Their report highlighted how new types of data could be utilised to extend the speed and reach of measurements with a view to ensuring that "No one should be invisible" (IEAG 2014, p.3). However, the International Union for the Scientific Study of Populations (IUSSP) warned that data quality issues should not be overlooked in the rush to measure an increasing number of indicators. They emphasised that national vital statistics and census data are essential ingredients in ensuring accurate measurement (IUSSP 2015, p.175).

Indeed, census data provide the backdrop against which most public service planning happens, including in South Africa. This is because the population and household counts from a census are crucial factors to establish the scale of any intervention required. The census also collects a variety of service-related information that is used for planning purposes, such as access to electricity. In fact, sustainable development goal 7 is to "ensure access to affordable, reliable, sustainable and modern energy for all." Indicator 7.1.2 is the "proportion of the population with primary reliance on clean fuels and technology." The United Nations' handbook on the measurement of the indicators notes that "Data for this indicator can be routinely collected at the national levels in most countries using censuses and household surveys" (UNSD,2021, indicator 7.1.2).

While censuses provide the necessary context for the service-level information, they are not conducted frequently enough to give the fine-grained information that policymakers require. Furthermore, the scale and the scope of census data collection imply that certain types of measurement problems are inevitable. In the case of South Africa, population undercount has been a repeated problem, as shown by the various post-enumeration surveys. To get a richer understanding of population and service-provision dynamics requires triangulating census information with other types of data, such as those emanating from social surveys or demographic surveillance sites.

Health and Demographic Surveillance System (HDSS) data is collected in many sites in Africa, South and Southeast Asia. These sites collect data on demographic and health characteristics of a population living in a well-defined geographic area (Sankoh & Byass, 2012). In South Africa, investment from the Department of Science and Innovation has enabled the harmonisation of the three existing rural-based HDSS sites, in Mpumalanga, Limpopo and KwaZulu-Natal provinces and the addition of urban-based nodes in Gauteng and the Western Cape (Herbst *et al* 2021) to create a publicly funded platform called the South African Population Research Infrastructure Network (SAPRIN) (Collinson *et al* 2022).

Information about the characteristics of the populations in these health and demographic surveillance sites (HDSS) is collected periodically, providing a rich source of data that can complement national censuses and surveys.

In this article we will examine some of the complexities that arise when one tries to measure access to electricity, as required for SDG 7. We will compare measurements on the electricity roll-out in a small area of South Africa from two sources: the SAMRC/Wits Agincourt health and demographic surveillance system (HDSS) site, and the national census. We describe the data sources in more detail below, but the important point is that we have two independent measures at the same points in time. Each source has its own strengths and weaknesses: the HDSS data is collected at a higher frequency and is subject to constant cross-checking in a way that is impossible with the census; but the census has much broader national coverage. We examine the patterns of electricity access as seen in consecutive national censuses, 1996, 2001, and 2011 in this area. We compare the number of households and households using electricity from the Agincourt HDSS at the village level as well as rough estimates of the population age-sex structure.

2. Literature review

2.1. Electrification

After the end of apartheid, South Africa's ex-"homeland" areas saw rapid electrification everywhere (e.g. Bekker *et al* 2008a, Gaunt 2005) which had implications beyond merely access to electricity (Dinkelman 2011). Progress on electrification in the late 1990s was so rapid that President Mbeki announced in 2004 that the electricity backlog should be eliminated by 2012 (Bekker *et al* 2008b, p.4). However, as Bekker *et al* (2008b) discuss, the meaning of this target and the accurate measurement of the state of electrification was not at all straightforward. The number of new connections, as well as an estimate of the total, was tracked by the National Electricity Regulator[5] (NER) up to 2003 whereafter the Department of Minerals and Energy (DME) took over that function, with a loss of granularity in the data and, in some cases, contradictory numbers.

Furthermore, the primary issue of interest to policy makers, as expressed for instance in the SDGs, is the percent of the population (or of households) that have access to electricity not the number of connections in themselves. To get an electrification rate, Eskom and the NER divided the estimate of the total number of connections by estimates of the number of households. However, the household counts available (for non-census years) were somewhat dubious. With measurement issues in both the numerator (the number of connections) and the denominator (the number of households), the DME switched to reporting estimates based on Stats SA survey data, where the electrification rate was measured by the fraction of households that used electricity for lighting (Bekker *et al* 2008b, p.9). As Bekker *et al* (2008b) note, the survey data produced a considerably higher electrification rate than the earlier calculations. Indeed in 2001 the estimates range from 55% of all households electrified to 70% (Bekker *et al* 2008b, Figure 5, p.9).

There may be good reasons for a mismatch between the total number of connections and the total number of households with access, since households sharing a dwelling may also share electricity access. This is on top of the problem of illegal connections which Bekker *et al* (2008b, p.9) also note. So, the number of connections may underestimate access. On the other hand, the total "stock" of connections as tracked by the NER also ignored the problem of disconnections (Bekker *et al, 2008b, p.9).*

The data from the Agincourt HDSS has already been used to analyse some of these issues. For example, Harris *et al.* (2017) used data to investigate the dynamics of household electricity connections and observed a net increase in the number of households that have access to electricity, changing from 69% in 2001 to 96% in 2013. They noted that in this aggregate increase, there are also short-term deviations due to: (a) new households forming faster than electricity connections, (b) people leaving connected households to set up new households in locations where there is no electricity access and (c) households that gain and lose electricity access over time. Wittenberg *et al.* (2017) used various data sources to decompose changes in household electricity access and found that electricity connections in households in the Agincourt HDSS increased by about 72% between 2001-2011.

While previous work using the HDSS data has broadly illustrated national trends from local data, there has been no attempt to directly cross-check the data. Given that census and survey estimates have effectively supplanted the type of "connection" data that tracked the electricity roll-out in the early years, it is important to get a sense of how accurate those estimates are, particularly given the wide range of possible electrification rates reported by Bekker *et al* (2008b). The HDSS data again provide a good opportunity for this sort of check, given that the measurement processes are largely independent of each other. Furthermore, the HDSS data is at a higher frequency and because of its longitudinal nature, there is more scope for picking up fieldwork errors and correcting them.

In this paper we will compare the electricity access data, and household and population counts from the Agincourt HDSS and the national census, to get a sense of what "noise" there might be in the

[5] Afterwards National Energy Regulator of South Africa (NERSA)

measurements. This should help to improve the way in which the census data is used as a benchmark for national planning (Statistics South Africa, 2010).

2.2. Population census count adjustment

A national population census aims to enumerate the whole population within the boundaries of a country. However, this is not always the case, at least for South Africa, as the census enumeration usually gets coverage errors. As a result, the realised census count needs to be adjusted to get a true estimate of the population. To derive the adjustment factors, called weights, the South African national statistics office, Statistics South Africa (Stats SA), conducts a post-enumeration survey (PES) immediately after the census. One of the objectives of the PES is to evaluate the completeness of the census data and provide statistical information on coverage errors. The PES is akin to a mini census conducted on a sample of the enumeration areas (EAs), i.e., the smallest geographical unit assigned to a single person to enumerate during a census. The PES data is compared with the census data and the estimated undercount rates used to derive weights for adjusting the census count. A detailed description of the methods used for each post-1994 South African census can be found in census documentation (Statistics South Africa, 2004; Statistics South Africa, 2012; Statistics South Africa, 1998).

Undercount rates are determined at a higher level of geographic unit and adjusted for other demographic variables to derive adjustment factors for lower geographic units. However, there are likely to be differential undercount rates for different areas. As a result, some areas may be over adjusted and others under adjusted especially at the local level. We will refer to census estimates as weighted/adjusted if the adjustment factors are applied and raw/unadjusted if adjustment factors are not applied.

2.3. Census and HDSS data compared

Using data from the most recent South African population census conducted in 2011, Shoko *et al.* (2016) found that the weighted census overestimated the Agincourt HDSS population whereas the unweighted census undercounted the population. They attributed the difference between unweighted census and HDSS population to different household definitions. The national census tries to measure people who were present in the area on the reference day. This corresponds to a "*de facto"* definition of household membership. The demographic surveillance sites, however, allow for a looser "*de jure"* definition, where temporary migrants are kept on the household rosters. This allows a more nuanced monitoring of the population. The HDSS data does allow one to distinguish between residents permanently in the area and the temporary migrants, but even restricting the counts to permanent residents is not exactly equivalent to measuring who precisely was in the area on the reference date. Consequently, some mismatch is to be expected.

An additional source of mismatch could arise in the enumeration of households. Both the census and the Agincourt HDSS employ similar definitions, in which co-residency and sharing resources ("eating from a common pot") feature prominently. Nevertheless, in the case of the HDSS the relationships are built up over many local census rounds, so that the demarcation of who belongs to what household is clarified. In the case of the national census operation, the enumeration would be begun afresh. One might expect that the history of local enumerations would make it more likely that the census would come to similar classifications, but it is not guaranteed that they should align exactly. Indeed, the number of households recorded in the 2011 national census compared to the number in the Agincourt HDSS data also showed differences. Both the weighted and unweighted census overestimated the number of households recorded in HDSS population.

An observable difference that probably reflects the different methodologies is the number of under-five children. There were more children counted in the national census, according to Shoko *et al.* (2016). It is possible that this reflects an overcount because high levels of child mobility have been reported within the study population (Collinson, 2010). Some children move between households, usually within the same family, and the HDSS makes a point of recording each child once, locating it in the household where they spend the majority of time. When household members are familiar with some children who don't always live there, it is possible that these 'overlapping' members are reported in more than one de facto household.

While Shoko *et al* focussed on the 2011 census in the Agincourt area, Nyirenda (2006) compared interpolated population estimates for the Africa Centre HDSS in KwaZulu-Natal (KZN) from the 2001 South African census with surveillance data from the same site. The study found that the weighted 2001 census population was greater than the Africa Centre HDSS population whereas the un-weighted census generally closely matched the surveillance data (Nyirenda, 2006). The KZN province had one of the biggest undercounts in the 2001 census and those adjustment factors derived at the province level ended up affecting estimates at the local level. In the case of the Africa Centre HDSS area, the presence of the site might have made enumeration easier, resulting in "raw" counts of much higher quality than elsewhere. Nyirenda (2006) also compared changes in the interpolated census estimates for the Africa Centre HDSS between 1996 and 2001 and observed an implausibly high intercensal population growth that suggests large under-enumeration in the 1996 census.

3. Methods

3.1. Data

Agincourt HDSS

We use surveillance data from the Agincourt health and demographic surveillance system (Agincourt HDSS, 2022) to compare electrification and demographic measures with the South African population censuses. As explained above, the HDSS monitors demographic, social and health characteristics of a geographically defined population over time. The baseline data for the Agincourt HDSS was collected from 20 villages in 1992. After 2007 the geographic area was increased. Furthermore, some new villages were constructed, so that there are now 31 villages covered by the HDSS site. To maintain comparability, we restrict the data to those villages that lie inside the original HDSS boundaries since these are the only villages for which we have data at the times of the 1996 and 2001 national censuses. Besides the original 20 villages the geographic area also encompasses three RDP villages constructed during the study period. At the time of the 2001 census there were 21 villages and 23 by 2011.

We use the Agincourt Energy Panel version of the Agincourt HDSS data (Agincourt HDSS, 2022).[6] This is constructed by taking repeated annual snapshots of the full Agincourt HDSS data on 30 November of each year. This means that there is a potential slight time mismatch with the census data which was collected in October. The electrification data was collected by means of a separate household asset schedule which was fielded only every second year starting in 2001. As noted in the document accompanying the Agincourt Energy Panel (Wittenberg *et al* 2022), there were a few households in every year that did not complete the asset schedule, so our Agincourt electrification estimates are adjusted upwards using a set of household weights released with the panel to correct for the missing asset information.

South African national censuses

We used data from the South African national censuses whose collection dates coincide with the operation of the Agincourt HDSS. We therefore use data from the 1996, 2001 and 2011 censuses. For all censuses, we use data obtained at the enumeration area (EA) level, which is the smallest geographical sub-division assigned to an enumerator during a census.

3.2. Areal interpolation

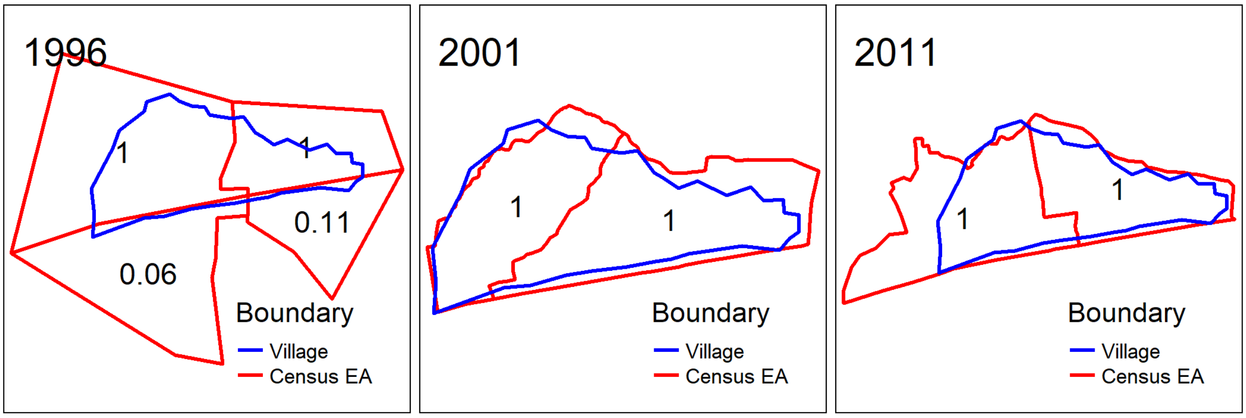
We want to obtain village level estimates of the number of households and households with access to electricity based on census EA estimates. However, the EAs and villages' boundaries do not line up. We therefore used areal interpolation techniques (Lam, 1983). In areal interpolation, we want to make estimates from a source polygon(s) to an overlapping, but incongruent target polygon(s) based on the overlapping area.

To calculate the proportion of the area that is overlapping, the polygons for the EAs and villages are first projected to an equal area projection. We then estimate the area for each EA and village. The two polygons are then superimposed to create a new polygon with intersecting areas. We calculate the area of

[6] We also use village identifiers, which are not available in the public release version.

intersection. The proportion of EA area overlapping with a village is then equal to the area of intersection divided by the area of the EA. However, some parts of EAs do not perfectly fall inside a village. We scaled the areal proportions of EAs that are overlapping such that they all sum to one. This allows us to re-apportion all households in an EA to at least one of the villages. We also scale and re-apportion intersecting areas of at least 1% of the EAs. This is done to avoid small intersecting areas of villages being apportioned more numbers from the overlapping EA. Figure 1 shows examples of the estimated scaled proportion of EAs overlapping with a selected village for the 1996, 2001 and 2011 censuses.

Figure 1: Proportion of EA area overlapping with village – 1996, 2001, 2011.



Assuming that we want an estimate value $\hat{y}_t$ of some attribute for the target polygon (village) t, then:

$$\hat{y}_t = \sum_{s=1}^{n} \frac{A_{st}}{A_s} * y_s$$

where A_{st} is the area of the zone of intersection between the source and target zones, A_s is the area of the source polygon and y_s is the value of the attribute for the source polygon (EA).

4. Results

4.1. Households

Our first set of results are contained in Table 1. It is evident that the raw counts in 2001 and 2011 agree remarkably with the HDSS counts. By contrast, there is a significant undercount of households in the 1996 census. These results are in striking agreement with Nyirenda's (2006) findings for Mtubatuba. The undercount in 1996 persists even if the census weights are applied. By contrast the 2001 and 2011 census weights over-inflate the household counts in the Agincourt area.

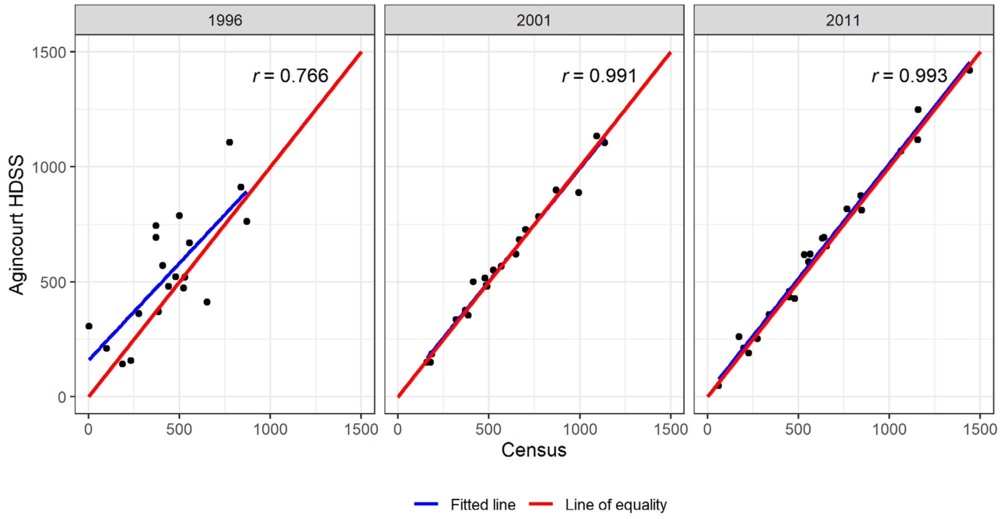
Table 1: Aggregate estimates of households – censuses vs HDSS

	Households			Percent error	
	Census		HDSS	Census vs HDSS	
Year	Raw	Weighted		Raw	Weighted
1996	8496	8931	10198	-16.7%	-12.4%
2001	11776	13287	11822	-0.4%	12.4%
2011	14079	16293	14433	-2.5%	12.9%

Household counts by village

In Figure 2 we show how the household counts from the HDSS stack up against the unweighted census counts. The red line is the line on which there would be 100% agreement between the national census counts and the HDSS ones. It is evident that the overall close match between the counts shown in Table 1 carries right down to the village level. The reported correlation coefficients are very close to one. It is also clear that the 1996 national census counts are much further from the HDSS ones. The fact that most villages are above the red line shows that in most villages there was a substantial undercount. Interestingly, however, in two villages there were substantial overcounts. The overall picture suggests that the 1996 national census data was subject to considerably more error and noise than might have been desirable.

Figure 2: Estimated number of households – raw censuses vs Agincourt HDSS.



4.2. Households with Electricity Access

Table 2 shows the counts of households with electricity access in the national censuses and the HDSS respectively. In this case we also need to make an adjustment for households in the HDSS that did not complete the asset schedule. Comparing the "raw" national census counts to the weighted HDSS ones, we again see an impressive level of agreement in 2001 and 2011. Unfortunately, there is no HDSS information about electricity access in 1996. It is uncertain how much error there might be in the 1996 census estimate of the percent electrified. If the undercount shown in Table 1 splits evenly across electrified and non-electrified households, then the percentage shown in Table 2 would still be accurate. Given what we know about the patterns of national electrification, the percentage shown in the table is likely to be in the right ballpark.

Table 2 Household electricity access – censuses vs HDSS

	Electricity for lighting				Percent electrified	
	Census		HDSS		Census	HDSS
Year	Raw	Weighted	Raw	Weighted[a]	Raw	Weighted
1996	2527	2654	-	-	29.7%	
2001	8045	9068	7579	8077	68.3%	68.3%
2011	13379	15476	11349	13371	95.0%	92.6%

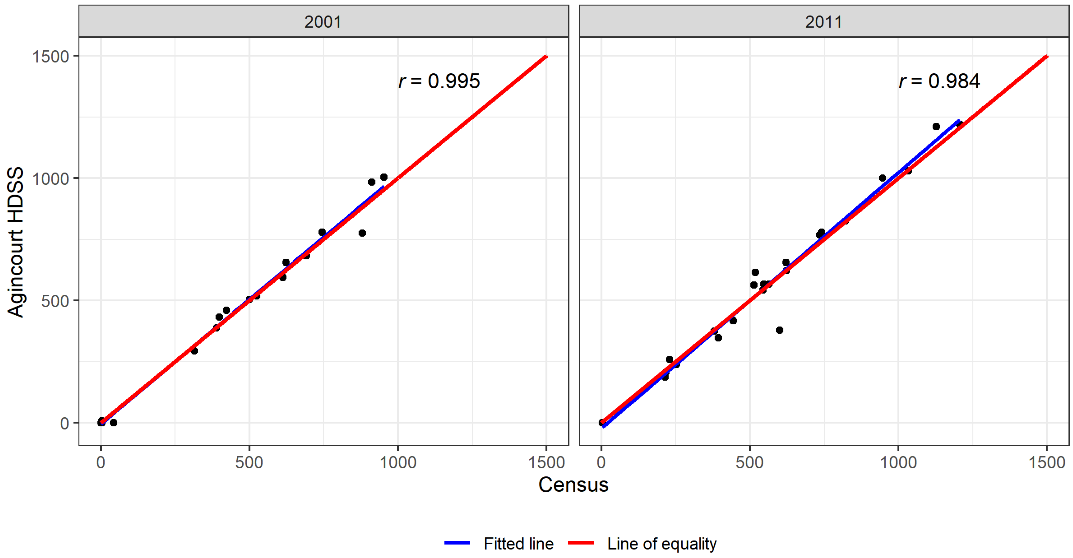
[a] The electricity data was adjusted to correct for households where the asset schedule was missing.

The electrification rates shown in Table 2 confirm the pattern of the national electricity roll-out, as discussed in the literature. The electrification rate in 2001 is much closer to the 70% that the survey evidence suggests for South Africa (nationally) than the 55% that was obtained by dividing the estimated "stock" of electricity connections by the number of households (Bekker et al 2008b). Interestingly also, the 2011 national census suggests a somewhat higher rate of electrification than the HDSS, suggesting that the slight undercount of households shown in the "raw" counts for 2011, may have disproportionately occurred among underserviced households.

Household electricity access by village

Again, we compare the electricity access numbers down to the village level. We do this comparing the unadjusted census counts to the weighted HDSS figures. The evidence is shown in Figure 3, where the line of equality (the red line) is again the important reference standard. Once more there are impressive levels of agreement as shown by the tight clustering around the line (with one or two exceptions) and the very high correlation coefficient.

Figure 3: Estimated number of households using electricity for lighting – raw censuses vs Agincourt HDSS.



Household access versus dwelling access

In the Agincourt HDSS we have not only household identifiers but also dwelling ones. It is possible to check therefore whether any mismatch between households and dwellings could account for the wide disparity in connection rates that was noted in the literature. As can be seen from Table 3, however, the count of dwellings closely matches the count of households, at least in the Agincourt area. In either census year not even half a percent of dwellings accommodated multiple households. This means that our results offer no explanation why household-based methods (e.g. using survey estimates) would lead to vastly different rates than monitoring of connections. This, of course, may be different in more urban settings.

Table 3 does show that there are mismatches. In those (few) cases where one dwelling accommodates multiple households there are many instances in which one of the households will have access to electricity while the other one does not. But the overall numbers are so low that it would not be advisable to draw conclusions from this evidence.

Table 3: Household counts versus dwelling counts in the Agincourt HDSS

Year	Households	Dwellings	Dwellings with multiple households	Mismatch[a]
2001	11822	11773	49	19
2011	14433	14380	44	23

[a] Mismatch records the count of dwellings in which one household had access to electricity and the other did not.

4.3. Population

From a policy standpoint, the number of households without electricity access is the key statistic, since this is a measure of the backlog. But from a welfare point of view the central issue is the number of people without access. Indeed, the indicators for SDG 7 are population and not household based. It is therefore of considerable importance to check how population-based measures compare.

In Table 4 we compare the unweighted census data against the Agincourt HDSS.[7] Results suggest that the 1996 census massively under-counted the HDSS population – the gap is 29% overall. The differences are larger for males aged 0-39 and across all female age groups. The unweighted 2001 census numbers closely match the Agincourt HDSS estimates, with the national census finding 1.7% fewer people. The unadjusted 2011 census coincides with the HDSS count, with a relative undercount of just 0.1%.

The age group specific errors are fascinating. In 2001 and 2011 there are relative overcounts among males aged forty years and older. Indeed, even in 1996 the relative rates suggest that the census found more men in this age group than in any of the other age-sex categories. One mechanism that might account for this is if the census picked up some absent migrants who were perhaps identified as household heads or other core members, even though they were not present in the area. This is likely to be due to a combination of respondent and fieldworker errors, e.g. household proxy respondents not remembering precisely who was living on the premises on the reference day, or due to fieldworker failure to insist on the reference period particularly as enumeration progressed. It is unlikely that there would have been an influx of migrants back into the area simply for purposes of being counted on census night.

Table 4: Unweighted census estimates and Agincourt HDSS

Year	Age group	Male		Female		Percent error	
		Census	AHDSS	Census	AHDSS	Male	Female
1996	0-19	12563	17776	12689	17615	-29	-28
1996	20-39	3915	5903	6152	8977	-34	-31
1996	40-59	1347	1540	2486	3565	-13	-30
1996	60+	1018	1097	1839	2415	-7	-24
2001	0-19	16569	17047	16683	16940	-3	-2
2001	20-39	5245	5329	8221	8464	-2	-3
2001	40-59	1844	1705	3630	3692	8	-2
2001	60+	1360	1335	2653	2651	2	0
2011	0-19	15822	15947	15753	15861	-1	-1
2011	20-39	6381	6391	9226	9517	0	-3
2011	40-59	2527	2128	4742	4679	19	1
2011	60+	1546	1439	3196	3190	7	0

4.4. People with electricity access

South Africa's reporting on electricity access has focussed either on the number of connections (in the late 1990s and early 2000s) or on the fraction of households with access (since then). That makes sense since infrastructure supply operates to dwellings or households. Nevertheless, SDG 7 is framed in terms of population access. We can again compare the census counts against those from the Agincourt HDSS. The information is shown in Table 5, which again shows excellent agreements between the census and the AHDSS figures. We have again used the unweighted national census counts. However, we applied

[7] Since the household counts match the HDSS much better with no weights, there seems to be no good reason for applying weights to the population counts. The impact of the weights would be to overestimate the 2001 population by 10.3%, while the 2011 weighted population count would be overestimated by 12.3%. The 1996 population would still be massively undercounted, by 21.2% overall.

weights to the Agincourt numbers to account for the missing asset schedules. The weighted counts are 4.5% and 9.2% higher in than the unweighted ones in 2001 and 2011 respectively.

Table 5 Number of people with access to electricity, by gender

	Census Unweighted	AHDSS Weighted	Percent error
1996			
Male	5874		
Female	7141		
2001			
Male	17876	18042	-0.9
Female	21876	22213	-1.5
2011			
Male	25188	24889	1.2
Female	31519	32264	-2.3

In Table 6 we re-express the counts of people with access as percentages of the overall population. As in the case of household connections, we see that the census has a small tendency to overestimate connections. By comparing these percentages to those in Table 2 we also see that population access is somewhat higher than household access, which is not surprising given that one household connection can supply many people. This does not address the question whether the type of connections available (mostly 20 amp) are equally good for supplying the needs of large households as small ones.

Table 6 Percentage of people with access to electricity, by gender

	Census Unweighted	AHDSS Weighted	Percentage point difference
1996			
Male	31.2		
Female	30.8		
2001			
Male	71.5	70.9	0.6
Female	70.1	69.8	0.3
2011			
Male	95.9	95.1	0.7
Female	95.8	95.3	0.5

5. Discussion

We wanted to explore whether census (and by implication survey) evidence is reliable for measuring access to services, as required for SDG monitoring. Our empirical results unequivocally indicate that the censuses provide good local level information – at least in an area that is under constant monitoring. Nevertheless, we have also found some unexpected patterns in the data. Our findings can be summarised as follows:

a) The national census and the HDSS agree better on household counts, population structure, and electricity access in 2001 and 2011 than one might have expected,
b) There seems to have been a massive undercount of households and people in the 1996 census,

c) The "undercount" weights from the census have the effect of overstating the numbers in well-measured localities; they are likely to understate them in badly measured ones,
d) Tracking electricity access by survey methods seems reasonable,
e) There seems to be some misreporting of the presence of older men. This suggests that population measures may be a bit more noisy than household ones, given that household membership is a tricky concept.

5.1. Agreement between the national census and the HDSS on households and electricity access

The level of agreement between the 2001 and 2011 national censuses and the HDSS data shown in sections 4.1 and 4.2 are better than one might have expected, given the vagaries of fieldwork, differences in the precise timing of the data gathering exercises, different processes of post-field quality controls and the possibility of different household respondents being interviewed with different opinions about how the household operates. The difference of 2.4 percentage points in the 2011 electrification rate (shown in Table 2) is, of course, not completely trivial but it is sufficiently small that the aggregate picture of electrification is reliable. It is encouraging that this is largely true even down to village level.

Interestingly, the agreement for 2011 according to our figures is **better** than for the counts reported by Shoko *et al* (2016) also for the Agincourt district. One of the major differences is that we restricted our analysis to the original (1992) study area, whereas Shoko *et al* compared the counts within the 2011 study area boundaries. The HDSS site was significantly enlarged in the late 2000s and by 2011 incorporated a peri-urban commuter belt in the north. It is plausible that enumeration may have been better in the parts of the field site that have been in the HDSS for longer.

It is also true that aligning the national census numbers with the AHDSS data is not trivial, given that the national census is working with its own spatial framework. Matching problems across the two geographies can reflect in differences in numbers.

5.2. There seems to be a problem with the 1996 census counts

The magnitude of the gap raises concerns. The household undercount is 17% on the "raw" counts and the population undercount is 28.7%. This does not seem to be an Agincourt specific problem. Nyirenda (2006) claims that the 1996 population counts in the Mtubatuba area could not plausibly be true given the 2001 HDSS counts and the 2001 census data for that area. To the extent to which these problems are generalisable, it suggests that some of the initial scepticism about the 1996 counts may have been warranted. The prevailing demographic models suggested that the census may have missed up to four million people. This was vigorously disputed by Orkin, Hirschowitz and Lehohla (1997), who suggested that the previous demographic information and the models in use at that time were inadequate.

5.3. The census weights will over- or under-adjust in different localities differently

In both Agincourt (this study) and in Mtubatuba (Nyirenda 2006) the unadjusted counts provide more reasonable estimates for the post-1996 censuses than the weighted ones. The main reason, as discussed also by Nyirenda, is that the weights derived from the post-enumeration surveys are calculated for much bigger geographic areas and then applied to local areas depending on their demographic structure. In areas where the measurements were conducted well (such as the demographic surveillance sites) the result is to over-inflate the counts, whereas in areas where the fieldwork was much poorer the adjustment will not be big enough. It is reassuring, though not surprising, that the quality of the enumeration in the demographic surveillance sites seems to have been better than elsewhere.

Unfortunately, the implication of our finding is that the (weighted) census counts will be less accurate at very small scales than is currently believed. The figures in Table 1 suggest that the margin of error in household counts may be of the order of ten percentage points. And in the case of population counts it is likely to be even bigger.

5.4. Tracking electrification by survey methods is a reasonable approach

We began our research with the puzzle posed by the findings of Bekker *et al* (2008b) that tracking electrification through survey methods produced higher levels of electricity availability than was the case by incrementing up the new connections that got logged by Eskom and municipalities. At least for our study site it seems that the connection rates are in the same ballpark as the ones documented nationally.

That suggests that survey measurement is reasonable. This may, of course, be different in urban settings where multi-household occupancy of dwellings may lead to much more complicated access and service issues.

5.5. Population-based estimates will be noisier than household-based ones

Our HDSS population counts show excellent agreement with the census numbers except for men aged forty years and older. This suggests that at least some absent migrants may have been put onto household rosters erroneously. More generally, counting dwellings and households in rural settings will be easier than pinning down household membership or, in the case of census enumeration, remembering who precisely was living on the premises on the reference day.

As noted before, from a welfare perspective it is the population-based measures which are the important ones and the ones which should be captured for SDG reporting. Our results are encouraging on that score, suggesting that despite some noise in the measurement of the population during the census, the overall the rate of access measure agrees with the AHDSS, give or take half a percentage point. Of course the errors in non-demographic surveillance sites are likely to be larger.

6. Conclusion

Our discussion has highlighted many issues which ought to be of interest to public policy planners. Besides the substantive points listed above, this study has again highlighted how important it is to measure socio-economic constructs from different perspectives. Triangulating information from the national census and national surveys against the data in demographic surveillance sites improves our understanding of the quality of the measurement processes in each.

References

Agincourt Health and Demographic Surveillance System. 2022. Agincourt Household Energy Panel [dataset]. Version 1.0: 1992-2015. Acornhoek: MRC/Wits Agincourt Unit [data producer]. Cape Town: DataFirst [panel producer]. Cape Town: DataFirst [distributor].

Bekker,B., Eberhard, A., Gaunt, T. & Marquard, A. 2008a. South Africa's rapid electrification programme: Policy, institutional, planning, financing and technical innovations. *Energy Policy*. 36, 3125-3137.

Bekker, B., Gaunt, C.T., Eberhard, A. and Marquard, A. 2008b. Uncertainties within South Africa's goal of universal access to electricity by 2012. *Journal of Energy in Southern Africa*, *19*(2), 4-13.

Collinson, M.A. 2010. Striving against adversity: The dynamics of migration, health and poverty in rural South Africa. *Global Health Action*. 3: 5080.

Collinson M.A., Mudzana T., Mutevedzi T., Kahn K., Maimela E., Gómez-Olivé F.X., Mngomezulu T., Gareta D., Kabudula C.W., Nemuramba R., Tlouyamma J., Gordon R., Tollman S.M., Herbst A.J. 2022. Cohort Profile: South African Population Research Infrastructure Network (SAPRIN). International journal of epidemiology. e206–e216

Dinkelman, T. 2011. The Effects of Rural Electrification on Employment: New Evidence from South Africa. *American Economic Review*. 101(7), 3078-3108.

Gaunt, T. 2005. Meeting electrification's social objectives in South Africa, and implications for developing countries. *Energy Policy*. 33, 1309-1317.

Harris, T., Collinson, M., & Wittenberg, M. 2017. Aiming for a moving target: the dynamics of household electricity connections in a developing context. *World Development, 97*, 14-26

Herbst, A.J., Juvekar S., Jasseh M., Berhane Y., Chuc N.T.K., Seeley J., Sankoh O., Clark S.J., Collinson, M.A. 2021. Health and demographic surveillance systems, history, state of the art and future prospects. Global Health Action, 14:sup1.

Independent Expert Advisory Group. 2014. A World that counts: Mobilising the data revolution for sustainable development. Available at www.undatarevolution.org.

International Union for the Scientific Study of Populations 2015. The IUSSP on a data revolution for development. *Population and Development Review*, 41(1),172-177.

Lam, N. S. 1983. Spatial interpolation methods: a review. *The American Cartographer, 10*(2), 129-150.

Nyirenda, M. 2006. *Small area projections: modelling the demographic and epidemiological dynamics of a rural area in South Africa.* University of Cape Town (Unpublished MPhil thesis).

Orkin, M., Hirschowitz, R., & Lehohla, P 1997. Accounting for the missing millions: The preliminary estimates from Census '96. *Indicator SA, 14(3),* 56-62.

Sankoh, O., & Byass, P. 2012. The INDEPTH Network: filling vital gaps in global epidemiology. *International Journal of Epidemiology, 41*(3), 579-588.

Shoko, M., Collinson, M. A., Lefakane, L., Kahn, K., & Tollman, S. M. 2016. What can we learn about South African households by comparing the national Census 2011 with the Agincourt Health and Demographic Surveillance System data in rural northeast Mpumalanga? *African Population Studies, 30*(2), 2403-2412.

Statistics South Africa. 1998. *Calculating the undercount in Census '96.* Pretoria: Statistics South Africa.

Statistics South Africa. 2004. *Post-enumeration survey: Results and methodology.* Pretoria: Statistics South Africa.

Statistics South Africa. 2010. *Population census 2011: strategy document - 20100118-v.4.3.* Pretoria: Statistics South Africa.

Statistics South Africa. 2012. *Post-enumeration survey: Results and methodology.* Pretoria: Statistics South Africa.

United Nations Statistics Division. 2021. *e-Handbook on SDG indicators*. Available at https://unstats.un.org/wiki/display/SDGeHandbook. Accessed 21 December 2021.

Wittenberg, M., Collinson, M., & & Harris, T. 2017. Decomposing changes in household measures: Household size and services in South Africa, 1994–2012. *Demographic research, 37*(39), 1297-1326.

Wittenberg, M., Collinson, M.A., Dinkelman, T., Kabudula, C., Machemedze, T., Twine, W., Kahn, K. and Tollman, S. 2022. *Agincourt Household Energy Panel User Guide*. Version 1. Cape Town: DataFirst.

About DataFirst

DataFirst is a research data service dedicated to giving open access to data from South Africa and other African countries.

We promote high quality research by providing the essential Open Research Data infrastructure for discovering and accessing data and by developing skills among prospective users, particularly in South Africa.

We undertake research on the quality and usability of national data and encourage data usage and data sharing.

DataFirst

www.datafirst.uct.ac.za
Level 3, School of Economics Building,
Middle Campus, University of Cape Town
Tel: +27 (0)21 650 5708
info@data1st.org / support@data1st.org

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD