What can we learn about South African households by comparing the national Census 2011 with the Agincourt Health and Demographic Surveillance System in the rural northeast Mpumalanga?

Shoko, M[1]; Collinson, M[2]; Lefakane, L[1], Kahn, K[2]

Affiliation: [1]Statistics South Africa, [2]University of the Witwatersrand

Abstract

Households are central to the demographic processes; fertility, mortality and migration. While the definition of a household is not in dispute across surveys in the country i.e. it is a socio-economic unit based on shared resources and co-residency; what constitute a household differs by enumeration methodology. Researchers have, however, not fully taken advantage of the availability of the multiple data sources for measuring demographic phenomena. Among the sources of household data in the country are Health and Demographic Surveillance Systems (HDSS) and the national population and housing census data. This study uses the cross-sectional national population and housing census and the longitudinal Agincourt HDSS data to examine the number of households, household size and population size in a rural sub-district. Key findings are that differences exist in household and population estimates between the national census and HDSS. This could be explained by high levels of temporary migration seen in the HDSS data but are less evident in census data. Also, the differences in children may be explained by local mobility evident in the census data and less evident in the HDSS data. Differing definitions and data collection procedures are the sources of differing population results, especially the number and size of rural households.

Introduction

Households are important socio-economic units and are at the core of service delivery driven by the 1998 South African National Population Policy that places population at the centre of policy and development, and regards people as the driving force and ultimate beneficiaries (South Africa 1998). The National Development Plan (NDP) stresses the importance of healthy and resilient households to achieve the country's vision for development. Keeping track of the number and size of households has its challenges.

Researchers have shown the average household size is coming down largely as a result of an increase in the proportion of smaller households being formed (Wittenberg and Collinson 2008). The number of households requiring services can be a moving target as the backlogs for service provision remain despite the progress made over the years partly due to the growing number of households. The question on household dynamics is therefore of policy relevance and evidence needs to be compiled.

As a wide variety of data becomes more accessible, the need for triangulation of different data sets becomes important and analysts must know the rules defining each dataset. In this paper we compare the Health and Demographic Surveillance (HDSS) Centre and the national census in 2011. The rules of HDSS and national Census data are slightly different and these enable the triangulation because the results reflect the same population in two different ways. The HDSS is embedded in a particular geographical area and conducts routine updates of vital statistics. Due to constant return to the household by the HDSS fieldworkers, the household memberships are regularly confirmed, which enables the HDSS operation to

maintain a more complex household definition, as explained below. The census is a national operation and provides the full cross-section of all residents on census night.

The objectives of this study are to compare data in the same geographic area from, the national census in 2011 and from HDSS records in the same year to gain insight into household dynamics that go further than can be provided from any one dataset. We will compare the number of households and population size; the average household size in the sub-population, and the population age-sex structure from the two data sources.

Data Sources

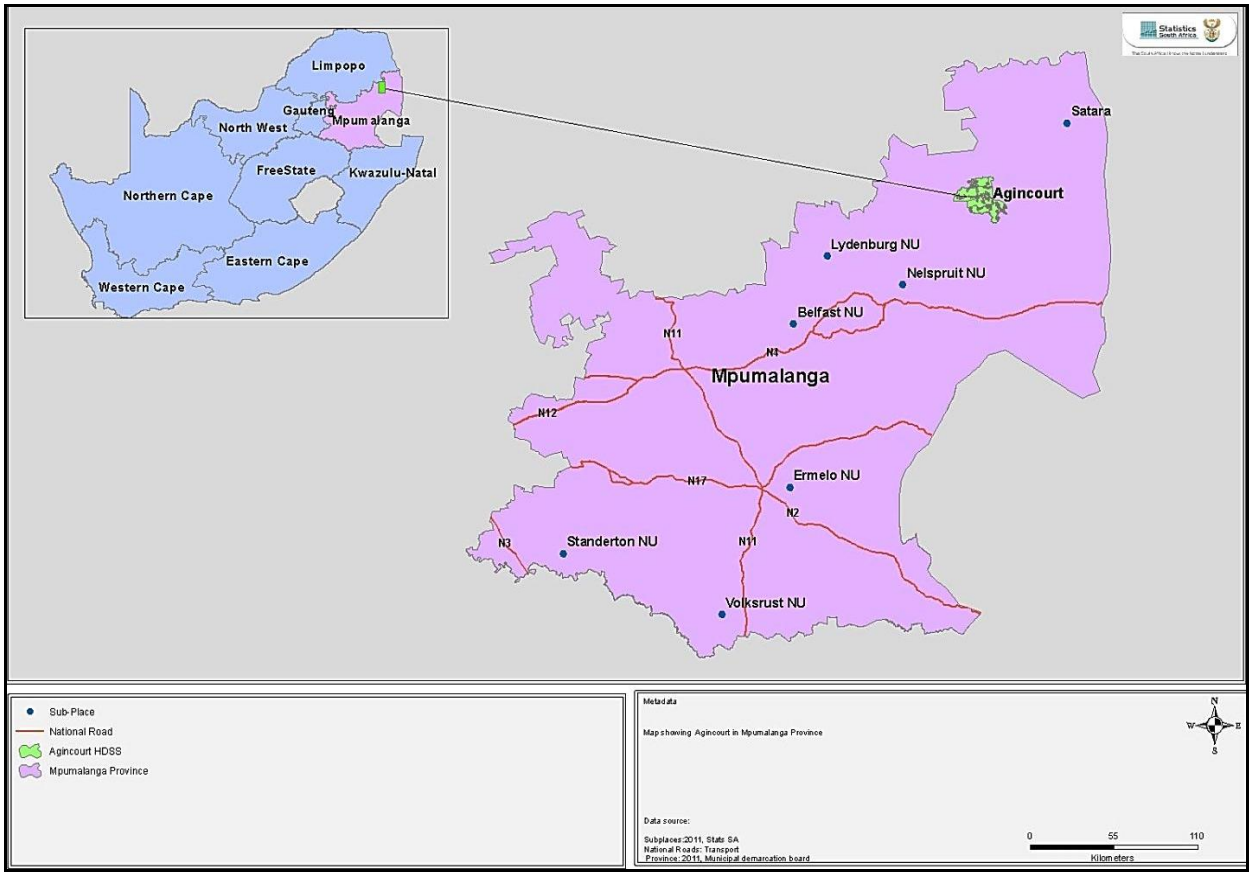
Agincourt Health and Demographic Surveillance System

The Agincourt Health and Demographic Surveillance System (HDSS) is based in a sub-district of Bushbuckridge situated in Mpumalanga province. It is located in the rural northeast part of the country close to the Mozambique border (Figure 1). Historically, this was part of the “homeland” of Gazankulu, characterised by poor agricultural land, underdeveloped infrastructure and maintained as labour reservoirs for the apartheid government (Kahn et al. 2012).The area has experienced a great deal of circular migration both before and after the fall of apartheid although with different patterns (Collinson, Tollman et al. 2006).

Agincourt HDSS is a member of the International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH), a network of HDSS Centres found in low and middle income countries. It is well-known and is the longest-running of the three Health and Demographic Surveillance Sites in South Africa (the others being Africa Centre HDSS and Dikgale HDSS). Agincourt HDSS is also diverse with

approximately one third of the population being self-settled Mozambican immigrants who were formerly refugees from the Mozambican civil war. The HDSS program commenced with a baseline census of the 21 villages in 1992 and later extended to incorporate six other villages in 2008 to currently covering an area of approximately 420km. From 1992 data has been collected longitudinally with an annual update of households' demographic and health data (Tollman 1999; Kahn, Collinson et al. 2012).

Figure 1: Location of Agincourt HDSS in relation to South Africa



South African population and housing census

The year 1994 is historic for South Africa because it marked its political transition from the apartheid epoch and its segregationist policies that among other things

resulted in the dearth of demographic information particularly for the majority of the African population to a democracy which commenced on the path to redress the structural inequality.

Censuses in South Africa date back to the first census in 1911, and thereafter, numerous other censuses were undertaken with the latest in 2011. According to Mostert, van Tonder and Hofmeyr (1987:3) cited in (Moultrie and Timaeus 2002), “the census coverage of the African population in the 1904, 1911 and 1921 censuses is viewed as poor in all respects, the 1936 and 1970 censuses as reasonably good, and those of 1946, 1951, 1960 and 1980 again as less good”. Post 1994 censuses conducted in 1996, 2001 and 2011; although with reported limitations of coverage and data quality issues; have provided widely available, comprehensive and nationally representative demographic data not available hitherto (Moultrie and Timaeus 2002).

Definitions used in the Agincourt HDSS

Like other South African HDSS, Agincourt HDSS has adapted the definition of households to suit the highly mobile populations in which they are based to include usual (permanent) members and members present for limited periods called temporary migrants (Kahn, Collinson et al. 2012; Hosegood Undated). The enumeration methodology is de-jure based and the definition of a household is the social group that usually resides and eats together, plus the linked temporary migrants who would eat with them on return. This definition retains links between temporary migrants and their rural household.

A temporary migrant is a household member who is away the majority of time, but retains a significant link to their base household. In analysis, a six month per year

cut-off point was chosen to differentiate ‘temporary migrants’ from ‘local residents’. Thus, people who are referred to as temporary migrants were absent from the household for more than six months of the year preceding observation, but who considered the index household in the HDSS to be their home base. Temporary migration status is based on ‘resident months’ status which records the amount of time each person is physically present in the household during the year preceding the census interview. The fieldworker hears the account of a person’s residence pattern and adds together the periods of home residence, rounds this up to a whole number and records it as the number of months that a person was present in the previous year. Based on the ‘resident months’, the fieldworker updates a ‘residence status’ variable, which has four categories, namely, ‘Local resident’, if ‘resident months’ is between six and twelve months; ‘Temporary Migrant’, if ‘resident months’ is less than six and the reason for absence is work-related; ‘Other Temporary Migrant’ if ‘resident months’ is less than six and the reason is not work-related; and ‘Visitor’ is if a person was present at the census but should not be considered part of the household.

The residence definition for recording children is to record them at the place where they spend the majority of their time, even if there is another household where a significant other adult stays, such as a separated parent.

The Agincourt definition of permanent migrant is a person who enters or leaves a household with a permanent intention. This includes people who leave the index household and establish a household or join a household elsewhere. The respondent informs the fieldworker that the migration is ‘final’ (Madhavan, Schatz et al. 2009). An implication of this definition is that when a field worker encounters a

permanent out-migrant the person is removed from the household list, whereas a temporary migrant is not.

Definitions used in the National Census

The South African population and housing censuses follow the de-facto methodology, whereby people are enumerated in reference to where they were at the census reference period regardless of being usual residents or visitors in the respective places. The census de-facto approach has the advantage of circumventing double counting at a national level, because it is clear who is resident in the household during the census reference period. The assumptions for de-facto population is that the majority of the population is enumerated at their place of usual residence which may sometimes be problematic for places like South Africa and the rest of sub-Saharan Africa where households can be complex, fluid and geographically dispersed (McDaniel and Zulu 1996; Young and Ansell 2003). The household questionnaire was one of the three questionnaire types available for census and is from which the household information for individuals in a household is obtained. The enumeration period was from the 10th to the 31st of October, but due to logistical constraints, enumeration was extended for more weeks but with the midnight of the 9th to the 10th of October as the reference date (Statistics South Africa 2012).

Census 2011 weighting methodology

Dual system estimation was used to arrive at the true population of the country. This means that two independent sources or ‘systems’ are used to arrive at the estimate of the true population: the census and the Post-enumeration survey (Wharf Higgins, Strange et al.). The PES is an independent sample survey that is conducted

immediately after the completion of census enumeration in order to evaluate the coverage and content errors of the census. The PES for Census 2011 was undertaken shortly after the completion of census enumeration. Households are therefore observed twice i.e. by PES and census. Both estimates contribute to the dual-system estimate of the true population. In the end, this true population is compared with the census-enumerated population and the difference is the net undercount (or overcount). From this, weights are then developed to adjust the census population to the true population. A detailed methodology can be found on Statistics South Africa (StatsSA) website: http://www.statssa.gov.za/.

Triangulating Census and Agincourt HDSS data

The study uses the household Agincourt HDSS data and the corresponding data from the 2011 South African population and housing census data for the same geographic area.

Several Agincourt HDSS field staff were involved in that geographical sub-district in census enumeration which is good for empowerment of local communities and to ensure community co-operation. We are not certain of the extent to which that would have affected compliance with census methodology.

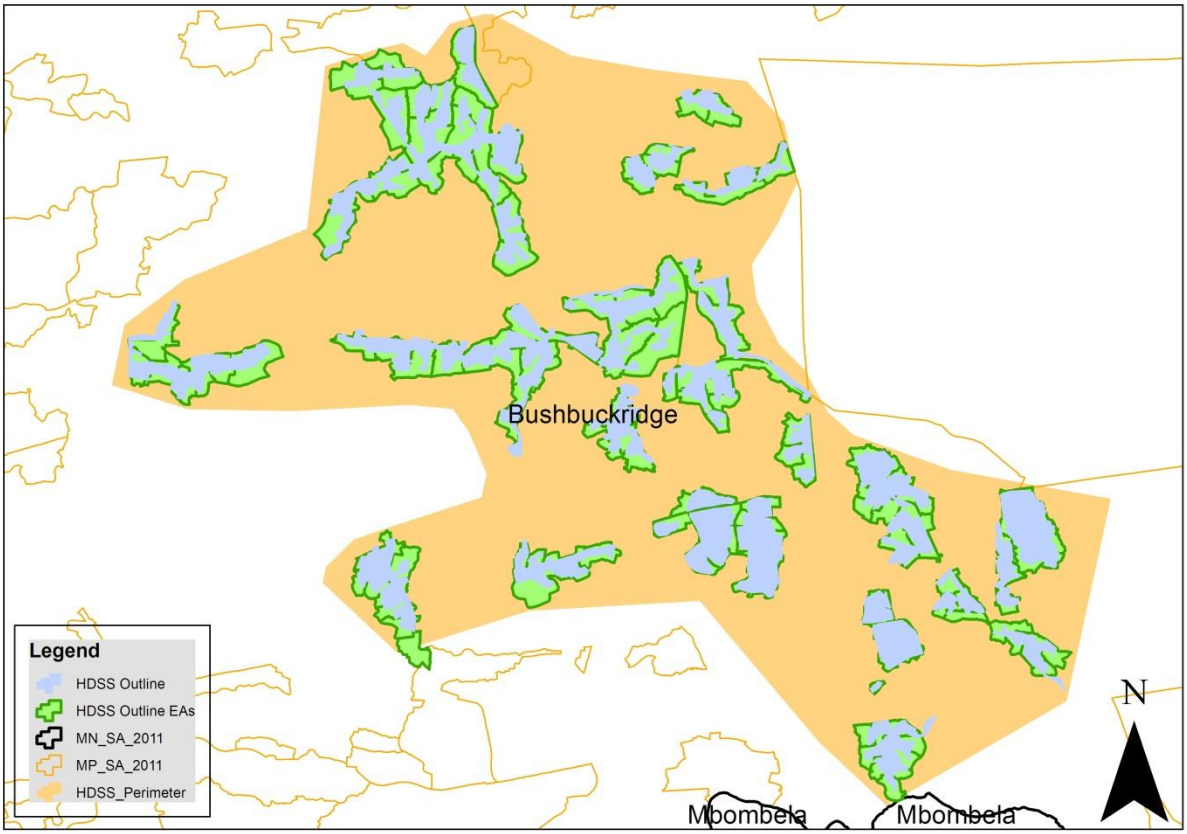
We applied the census weight variable for census estimates but clear variations have been noted in census coverage for sub-national levels and we are not certain of the coverage for Agincourt area specifically. In the study we therefore present both weighted and unweighted figures.

It is important to note that Agincourt population refers to the population on 1st July 2011 because this was the date on which the data was presented, while Census 2011 refers to the midnight of the 9th to the 10th of October, however because the difference in the periods is relatively short (about 3 months), the assumption is that this will not cause any significant difference in the data and therefore will not affect the comparison.

We used Agincourt HDSS ESRI shapefiles and census shapefiles and because the shapefiles came from two different sources; procedures had to be followed in order to have comparable data. Agincourt HDSS data includes ESRI shapefiles obtained from the global positioning system (GPS) coordinates obtained from the georeferencing of residential units using the geographic information system (GIS) technology. The perimeter and the outline shapefiles from Agincourt HDSS data show the geographical boundary for the site and for each of the villages respectively. Shapefiles exist at different census geography levels with the lowest being at Enumeration Areas (EAs). For logistical and administration of census undertaking, the country is divided into EAs which are the smallest geography units. The 2011 census shapefiles at municipal level come from the Municipal Demarcation Board of South Africa whose function of delimiting the country's district, municipality and electoral ward boundaries is provided for in the country's constitution. The board uses the World Geodetic System 1984 coordinate reference system and geographic datum. Statistics South Africa Geography department then further demarcates the shapefiles to lower geography levels.

The Agincourt HDSS shapefiles were overlaid on the census shapefiles. We then selected EAs that were intersected by the border of the HDSS perimeter. Overlay of Agincourt shapefiles as well as EAs from census 2011 with the perimeter boarder of Agincourt is shown in Figure 2 below. The map shows that the overlay of the two data sets was near perfect and therefore reasonably comparable geographically. We then extracted household data for the respective EAs from census data. We assume that Agincourt area typifies the South African rural and former homelands and so examination of the trend in Agincourt household structure and composition using HDSS data to some extent can be extrapolated to similar areas. Additionally; the corresponding analysis of household size can be understood in the context of “rural to urban” migration.

Figure 2: Agincourt HDSS using shapefiles from 2011 census and HDSS shape files

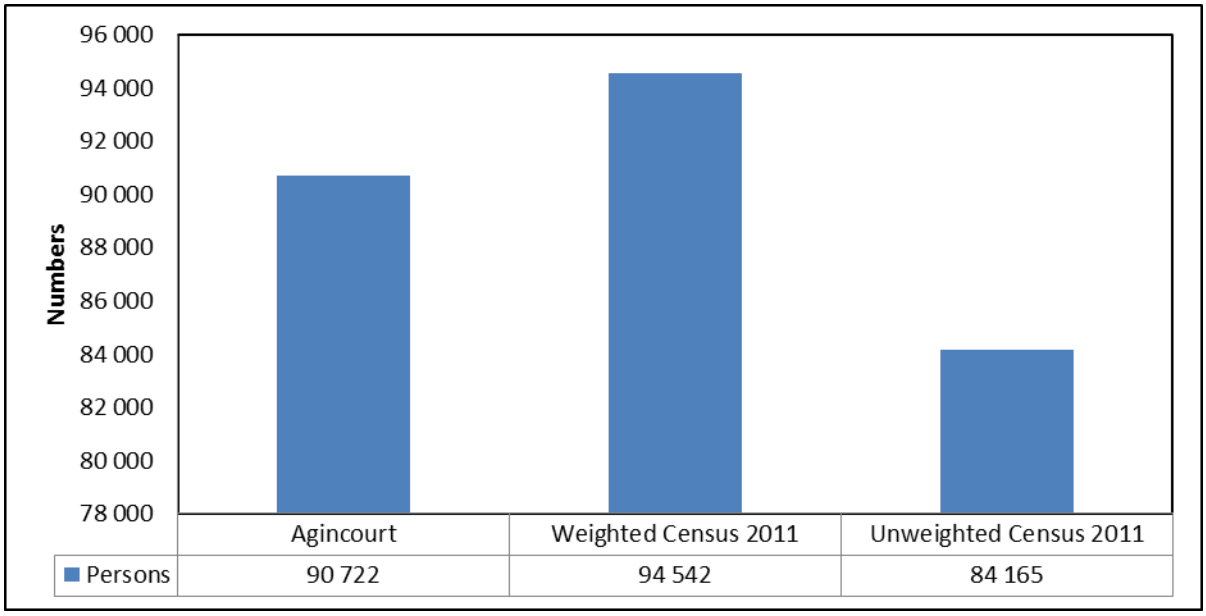


Results

The results are presented as comparable population indicators in 2011 from the national census and HDSS data from the same geographical area.

Figure 3 shows that the overall population size in the unweighted national census is 84,165 and the weighted number is 94,542. The HDSS gives the population number at 90,000 persons.

Figure 3: Number of persons by source of data



The number of households in the sub-population differ between the national census and HDSS (figure 4). The national census weighted data shows 23,000 households and unweighted 20,060 households. The Agincourt HDSS has 16,000 households in the same population.

Figure 4: Number of households by source of data

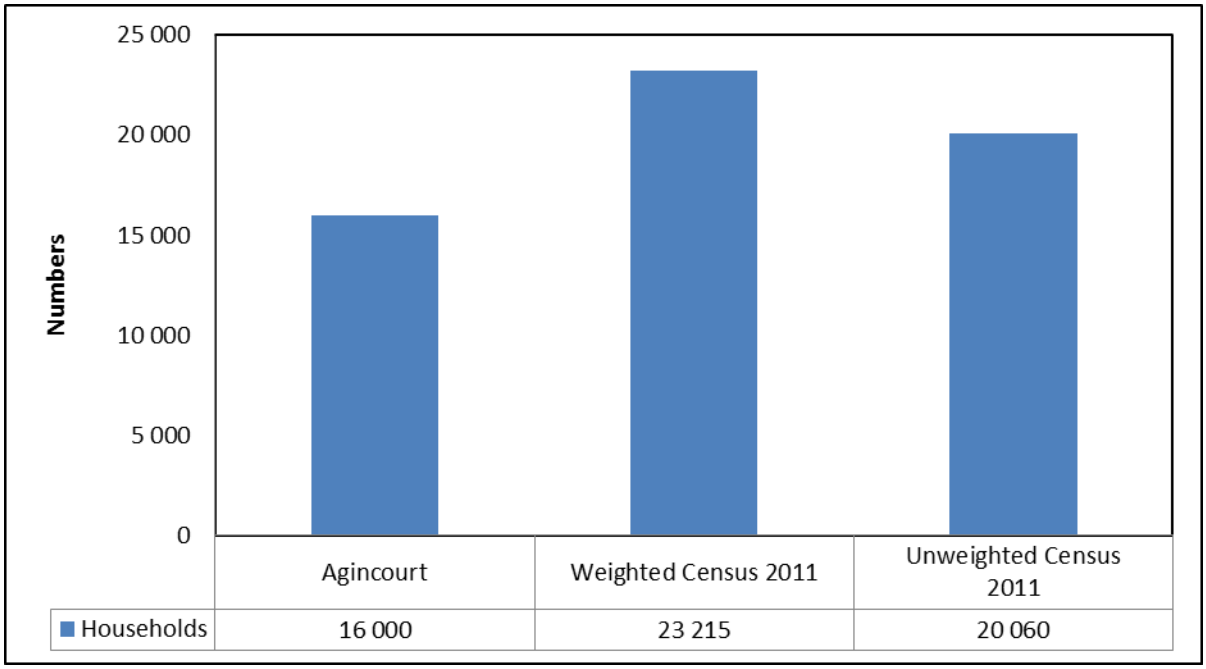
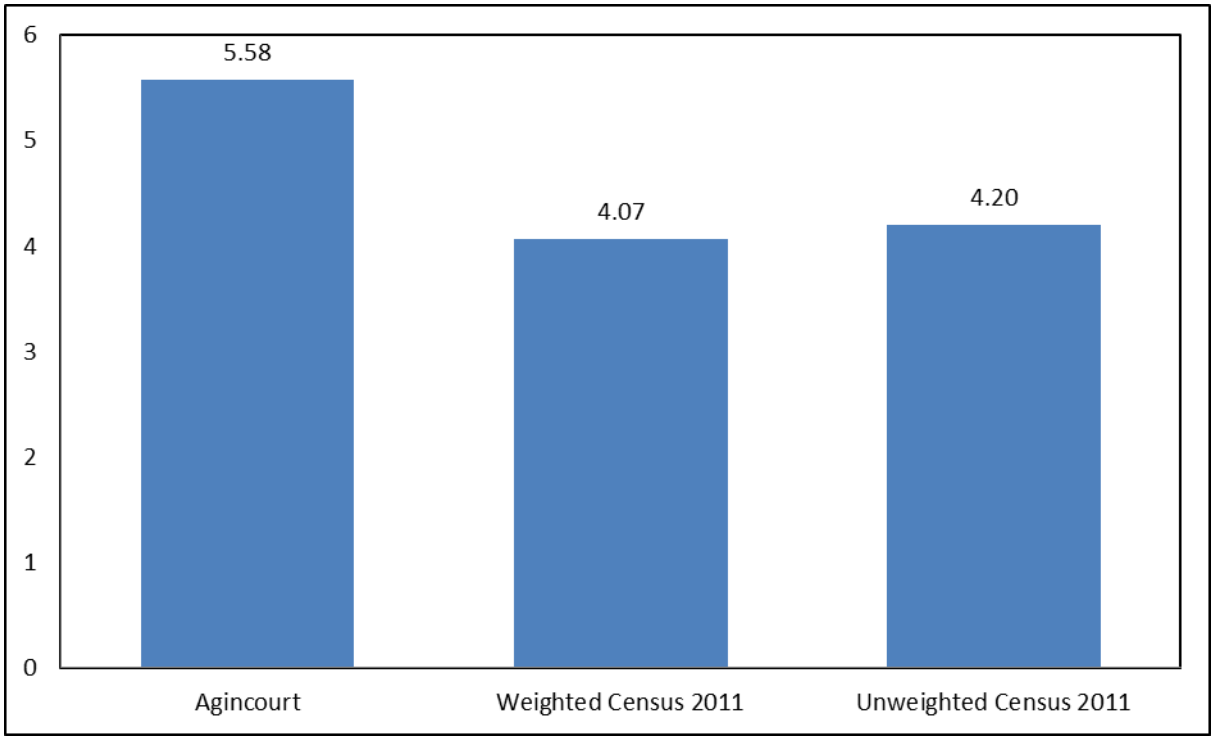


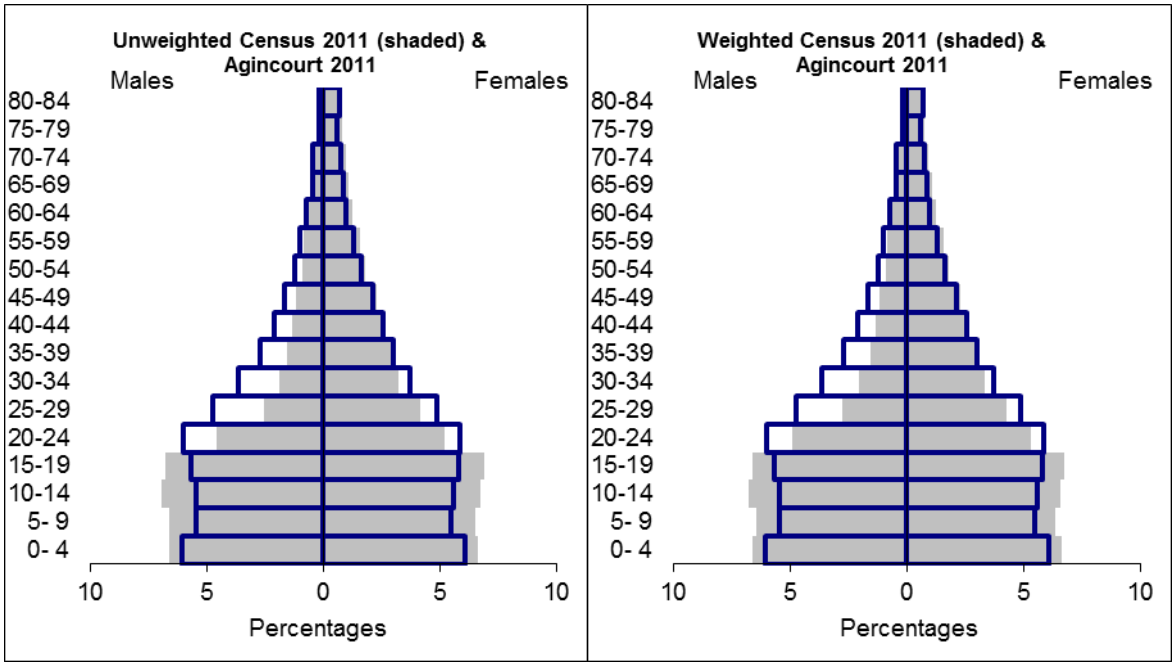
Figure 5 shows that the average household size in this sub-population is 5.58 for the HDSS in 2011, which is more than a person per household larger than in the national census. The national census has an average household size of 4.07 for the weighted data and 4.20 for the unweighted data.

Figure 5: Average household size by source of data



In figure 6 the age and sex structure of the population is compared from the two different data sources. For the age-groups 55 years and older, there is a good match of the number of persons in each age-band for both sexes. In the age-group 20-54 years the Agincourt data has higher counts of working age adult ages, especially for males, but also for females in the 20-34 age group. Conversely, in the 0-19 age group the national census records more children in the sub-population than HDSS. In each 5-year age-band under age 19 years the HDSS data has slightly fewer children recorded than in the national census. This is the opposite in the ages 0-19 years where difference is in the other direction and

Figure 6: Age-sex structure by source of data



Discussion

Each population census, including the national census and the HDSS, seeks to enumerate the full population, therefore an initial expectation when comparing two census methods for the same population is that the resulting numbers would be similar. The observed differences highlight the fact that for each method the definitions and data collection procedures are somehow reflected in the results.

Important dimensions of difference in national census and HDSS methodology are as follows. In household definitions, the HDSS has a *de jure* and the national census a *de facto* definition. In data collection procedures, the HDSS has a longitudinal approach and the national census is cross-sectional. In scale, the HDSS is confined to a sub-district of around 90,000 persons and the national census includes the whole population of around 52 million people, in 2011.

The purposes of data collection also vary slightly. The national dataset is a large administrative exercise required by the state to provide insight into population processes, denominators for planning and for policy-making. The HDSS is closer to the concept of a population laboratory, run by health and population scientists, to capture the dynamics in the population, assess mechanisms involved in population change and test specific interventions, also to inform policy-making.

Assessing the differences in population numbers described above the easiest difference to explain is the fact that HDSS households are on average larger in size. This is a direct result of the household definition, which deliberately includes the temporary migrants who belong to the rural household but are temporarily absent for work or education purposes. This innovation in household definition is made possible because households are repeatedly visited and the residence status of members can be verified. It is also made possible by the small geographic scale of data-collection. Double-counting can be avoided by reconciling migrations within the study site and the fact that most temporary migration occurs is to a destination outside the study site. The proportion of rural household members who are temporary migrants is high, reflecting the structural impact of Apartheid labour migration which forced African populations to reside far from employment opportunities. Rural populations still reside on land that is traditionally owned, which means that poorer households benefit from the low cost of the land and minimal rents. Hence, rural households still tend to be multi-local to access the potential benefits of cities, towns or commercial farms, while retaining a family base in a rural village.

The *de facto* household definition used by the national census, and emulated by national censuses around the world, employs a specific time reference to avoid counting an individual more than once. Also, temporary migrants linked to a

household are qualitatively self-defined by the household, which could produce an unsatisfactory imprecision that could result in an over-count. For example, if a labour migrant woman works in Johannesburg, but has a household in the rural area where her mother looks after her children; it would be difficult from a data perspective not to have the woman counted in her workplace residence as well as in the rural household.

In the absence of triangulation it may be difficult to see the extent of temporary labour migration at a population level. Using national census data alone would result in the population counts being biased towards places near work or education, while the vital role of rural (or peri-urban) households can be overlooked in the national socio-economy.

The comparison of age sex pyramids above re-emphasises the presence of high levels of temporary circular migration. The age groups affected by migration are precisely the ones that are larger in the HDSS and smaller in the national census.

The differences in the number of children captured in the national census and HDSS is interesting, and harder to explain. Here we see higher numbers of children on aggregate captured by the national census. This needs exploration to find the source of the difference and again we must look at the way data is collected. It is thought that children of all ages are locally mobile in the rural areas, especially in areas of poverty (Madhavan et al, 2012). Firstly, there is the possibility of a child's parents residing in different locations and the child moving between parental locations. Secondly, there is the practise of fostering, where children relocate to family members that can afford to look after them or who live closer to health services and schools. Within the HDSS, this is handled by a fieldworker determining which is the

main place of residence for a child, i.e. the place where they spend the most time. So the social reality of children's multi-local residence is simplified in order to model children's residence as accurately as possible from an exposure perspective, but not overcomplicate the surveillance operation.

The social reality bursts out of the HDSS household definition and there may be more local mobility of children than can be captured in the HDSS system. This is reflected in the national census data but with possible double-counting of children as more than one household claims the membership of a multi-local child. This is not following the census rules unless a child stays in two households in census week, but slippage of the household definition can be imagined, because people are aware that the census is used to plan public services and unless there is strict probing by the fieldworker a child can be listed in a household even if they are not residing there on census night.

The fact of locally mobile children is complex to manage in either system. As with the case of labour migration, it can be a direct consequence of the socially disruptive historical context and the subsequent spatial inequities that still plague the country. The fabric of home and work remain tenuous, especially in the poorest households and children are mobile as a result. The challenge of reflecting this in census data is shown by the triangulation used in this paper.

The difference in population numbers in the same geographic area reflected in the two systems is a consequence of the different household definitions. The HDSS should show a larger population, as it does in comparison with the unweighted data. On aggregate this is slightly off-set by the census picking up more children without which the difference would be even greater. The weighting of the national census

data recovers the difference and shows that the weighted number of the *de facto* population is closer to the de jure population shown in the HDSS.

The larger number of households recorded in the national census compared to the HDSS is a harder result to explain. There can be structural features contributing to this difference. In a recent community meeting in Agincourt a village elder said that when using a household definition of eating from the same pot this puts people together that otherwise live apart (Rhian Twine – personal communication). In other words, the census is picking up more households than the HDSS because households that live in separate dwellings while sharing resources report themselves as separate households during the census. More work is needed to explore whether this can account for the gap between the numbers of household counted in the national census and the HDSS.

Conclusion

It should not be surprising that different household definitions result in different household and population counts. It can be instructive to use these differences to learn some of the complex dynamics that exist in South Africa rural areas, particularly due to high levels of migration and local mobility.

These can be explored further by comparing the national census with HDSS data in different years, as well as repeating the exercise with other HDSS Centres in other parts of the country, namely the Africa Centre HDSS in KwaZulu-Natal and the Dikgale HDSS in Limpopo.

A systematic process of triangulation and interrogation of census results can provide insight into household dynamics, especially in highly mobile populations. Observed

differences can have implications for the training of census fieldworkers and for the interpretation of analytic results.

References

Collinson, M., S. Tollman, et al. (2006). "Highly prevalent circular migration: households, mobility and economic status in rural South Africa." Africa on the move: African migration and urbanisation in comparative perspective: 194-216.

Hosegood, V. (Undated). The Measurement of Migration in South African Demographic Surveillance Systems.

Kahn, K, MA Collinson, FX Gómez-Olivé, O Mokoena, R Twine, P Mee, SA Afolabi, BD Clark, CW Kabudula, A Khosa, S Khoza, MG Shabangu, B Silaule, JB Tibane, RG Wagner, ML Garenne, SJ Clark and SM Tollman (2012). "Profile: Agincourt Health and Socio-demographic Surveillance System." International Journal of Epidemiology 41(4): 988-1001.

Madhavan, S., E. Schatz, et al. (2009). "Effect of HIV/AIDS-related mortality on household dependency ratios in rural South Africa, 2000-2005." Population Studies 63(1): 37-51.

Madhavan, S., Schatz, E., Clark, S.J. Collinson, M.A. (2012) Child Mobility, Maternal Status and Household Composition in Rural South Africa. Demography 49, no. 2: 699-718.. doi: 10.1007/s13524-011-0087-3.

McDaniel, A and E Zulu (1996). "Mothers, Fathers, and Children: Regional Patterns in Child-Parent Residence in Sub-Saharan Africa." African Population Studies 11(1).
Moultrie, T and I Timaeus (2002). Trends in South African Fertility between 1970 and 1998: An Analysis of the 1996 Census and the 1998 Demographic and Health Survey Technical Report. Cape Town, Burden of Disease Research Unit: Medical Research Council.

South Africa. (1998). Population Policy for South Africa. Department of Social Development.

Statistics South Africa. (2012). Census 2011 Metadata. Pretoria: South Africa.

Tollman, S. (1999). "The Agincourt Field Site – Evolution and Current Status." South African Journal of Medicine (89): 855-857.

Wittenberg, M. and M. A. Collinson (2008). Restructuring of Households in Rural South Africa: Reflections on Average Household Size in the Agincourt sub-district, 1992–2003.

Young, L and N Ansell (2003). "Fluid Households, Complex Families: The Impacts of Children's Migration as a Response to HIV/AIDS in Southern Africa*." The

Professional Geographer 55(4): 464-476. What can we learn about South African Households from Post-1994 Census and Health and Demographic Surveillance System Data?