Technical notes on the data sources

General Household Survey[1]

The GHS is a multi-purpose annual survey conducted by the national statistical agency, Statistics South Africa (Stats SA), to collect information on a range of topics from households in the country's nine provinces. The survey uses a sample of 30,000 households. These are drawn from Census enumeration areas using a two-stage stratified design with probability proportional to size sampling of primary sampling units (PSUs) and systematic sampling of dwelling units from the sampled PSUs. The resulting weighted estimates are representative of all households in South Africa.

The GHS sample consists of households and does not cover other collective institutionalised living-quarters such as boarding schools, orphanages, students' hostels, old-age homes, hospitals, prisons, military barracks and workers' hostels. These exclusions should not have a noticeable impact on the findings in respect of children.

Changes in sample frame and stratification

The sample design for the GHS 2010 was based on a master sample that was originally designed for the Quarterly Labour Force Survey (QLFS) and was used for the GHS for the first time in 2008. The same master sample is shared by the GHS, the QLFS, the Living Conditions Survey and the Income and Expenditure Survey. The previous master sample for the GHS was used for the first time in 2004. This again differed from the master sample used in the first two years of the GHS: 2002 and 2003. Thus there have been three different sampling frames during the nine-year history of the annual GHS, with the changes occurring in 2004 and 2008. In addition, there have been changes in the method of stratification over the years. These changes could compromise comparability across iterations of the survey to some extent, although it is common practice to use the GHS for longitudinal monitoring and many of the official trend analyses are drawn from this survey.

Provincial boundary changes

Provincial boundary changes occurred between 2002 and 2007, and slightly affect the provincial populations. The sample and reporting are based on the old provincial boundaries as defined in 2001 and do not represent the new boundaries as defined in December 2005.

Weights

Person and household weights are provided by Stats SA and are applied in *Children Count – Abantwana Babalulekile* analyses to give estimates at the provincial and national levels.

Survey data are prone to sampling and reporting error. Some of the errors are difficult to estimate, while others can be identified. One way of checking for errors is by comparing the survey results with trusted estimates from elsewhere. Such a comparison can give an estimate of the robustness of the survey estimates. The GHS weights are derived from Stats SA's mid-year population estimates. For this project, weighted GHS population numbers were compared with population projections from the Actuarial Society of South Africa's ASSA2008 AIDS and Demographic model (full version), which is regarded as a "gold standard" for population estimates.

Analyses of the nine surveys from 2002 to 2010 suggest that over- and under-estimation may have occurred in the weighting process:

- When comparing the weighted 2002 data with the ASSA2008 AIDS and Demographic model estimates, it seems that the number of children was under-estimated by 5% overall. The most severe under-estimation is in the youngest age group (0 – 9 years) where the weighted numbers of boys and girls yield under-estimations of 15% and 16% respectively. The next age group (5 – 9 years) is also under-estimated for both boys and girls, at around 7% each. The difference is reduced in the 10 – 14-year age group, although boys are still under-estimated by around 1% and girls by 3%. In contrast, the weighted data yield over-estimates of boys and girls in the upper age group (15 – 17 years), with the GHS over-counting these children by about 5%. The pattern is consistent for both sexes, resulting in fairly equal male-to-female ratios of 1.02, 1.01, 1.03 and 1.01 for the four age groups respectively.
- Similarly in 2003, there was considerable under-estimation of the youngest age groups (0 – 4 years and 5 – 9 years) and over-estimation of the oldest age group (15 – 17 years). The pattern is consistent for both sexes. Children in the youngest age group are under-estimated by as much as 16%, with under-estimates for babies below two years in the range 19 – 30%. The results also show that the over-estimation of males in the 15 – 17-year age group (9%) is much more severe than the over-estimation for females in this age range (1.4%), resulting in a male-to-female ratio of 1.09 in this age group, compared with ratios around 1.02 in the younger age groups.
- In the 2004 results, all child age groups seem to have been under-estimated, with the under-estimate being more severe in the upper age group (15 – 17 years). This is the result of severe under-estimation in the number of girls, which outweighs the slight over-estimation of boys in all age groups. Girls are under-estimated by around 6%, 8%, 8% and 12% respectively for the four age bands, while over-estimation in the boys' age bands is in the range of 2 – 3%, with considerable variation in the individual years. This results in male-to-female ratios of 1.10, 1.11, 1.12 and 1.14 for the four age groups.
- In 2005, the GHS weights seem to have produced an over-estimate of the number of males and an under-estimate of the number of females within each five-year age group. The extent of under-estimation for girls (by 7% overall) exceeds that of the over-estimation for boys (at 2% overall). These patterns result in male-to-female ratios of 1.06, 1.13, 1.10 and 1.13 respectively for the four age groups covering children.
- The 2006 weighting process yields different patterns from other years when compared to population estimates for the same year derived from ASSA2008, in that it yielded an under-estimation of both females and males. The under-estimation of females is greatest in the 0 – 4 and 5 – 9-year age groups, while the under-estimation of males is in the range 3 – 10% in the 5 – 9 age group and 1 – 6% in the 10 – 14-year age group. This results in male-to-female ratios of 1.09, 0.99, 0.96 and 1.00 respectively for the four age groups covering children.
- The 2007 weighting process produced an over-estimation for boys and an under-estimation for girls. The under-estimation of females is in the range of 4 – 8% while the over-estimation for boys is in the range of 1 – 5%. This results in male-to-female ratios of 1.07, 1.06, 1.08 and 1.06 respectively for the four age groups covering children.

- In 2008, the GHS weighted population numbers when compared with ASSA2008 over-estimated the number of boys aged 10 and over, in the range of 3% for the 10 – 14 age group, and 8% for the 15 – 17 age group. The total weighted number of girls is similar to the ASSA population estimate for girls, but this belies an under-estimate of female babies below two years (by 7 – 8%), and an over-estimate of young teenage girls. The GHS 2008 suggests a male-to-female ratio of 1.03 for children aged 0 – 4 years, which is higher than that of the ASSA2008 model.
- A comparison of the GHS and ASSA for 2009 suggests a continuation of the general pattern from previous years, which is that GHS weights result in an under-estimation of children in the 0 – 4 age group (especially infants), and an over-estimate of older children. In 2009 the under-estimation in the 0 – 4 age group ranges up to 4% for boys and 5% for girls. In the 15 – 17 age group, the GHS-weighted data produce population numbers that are 7% higher than ASSA for boys, and 3% higher for girls. The male-to-female ratios in 2009 are in keeping with those in ASSA2008, with the exception of the 15 – 17 age group where the GHS-derived ratio is higher, at 1.08, compared to 1.00 in ASSA.
- In 2010, the GHS weights again produce an underestimation of children in the 0 – 4 age group and an over-estimate of children aged 15 – 17 years. For the middle age groups, and for the child age group as a whole, there is less than 1% difference in the estimates from the two sources. For the 0 – 4 age group the under-estimate is lower than previously, at 2%, but for the oldest age group there is an over-estimate of 5%. The male-to-female ratios are similar across the two sources, although the ratio is 1.00 for all but the 0 – 4 age group in ASSA as against 1.01 for the youngest age group in ASSA and for all age groups in the GHS.

The apparent discrepancies in the nine years of data may slightly affect the accuracy of the *Children Count – Abantwana Babalulekile* estimates. From 2005 to 2008, consistently distorted male-to-female ratios means that the total estimates for certain characteristics would be somewhat slanted toward the male pattern. This effect is reduced in 2009, where more even ratios are produced, in line with the modelled estimates. A similar slanting will occur where the pattern for 10 – 14-year-olds, for example, differs from that of other age groups. Furthermore, there are likely to be different patterns across population groups.

Disaggregation

Stats SA suggests caution when attempting to interpret data generated at low level disaggregation. The population estimates are benchmarked at the national level in terms of age, sex and population group while, at provincial level, benchmarking is by population group only. This could mean that estimates derived from any further disaggregation of the provincial data below the population group may not be robust enough.

Reporting error

Error may be present due to the methodology used, ie the questionnaire is administered to only one respondent in the household who is expected to provide information about all other members of the household. Not all respondents will have accurate information about all children in the household. In instances where the respondent did not or could not provide an answer, this was recorded as "unspecified" (no response) or "don't know" (the respondent stated that they didn't know the answer).

Labour Force Survey (LFS) and Quarterly Labour Force Survey (QLFS)[2]

The LFS and its recent replacement, the QLFS, are nationally representative surveys conducted by the national statistics agency Statistics South Africa to provide population data on labour market participation. The LFS was a bi-annual survey, conducted in March and September each year from 2000 to 2007. This was replaced in 2008 by the QLFS, which is conducted four times a year.

The sample frame is designed as a general purpose household survey frame and is used by a range of other surveys including the General Household Survey (GHS), the Living Conditions Survey and the Income and Expenditure Survey. The sample covers the non-institutional population (ie private dwellings only) with the exception that it also includes workers' hostels. The sample size is roughly 30,000 dwellings, yielding individual data on approximately 100,000 resident individuals.

The sample is drawn from Census enumeration areas using a stratified two-stage design with probability proportional to size sampling of primary sampling units (PSUs) and systematic sampling of dwelling units from the sampled PSUs. The resulting weighted estimates are representative of all households in South Africa, and are designed to be representative at provincial level and within provinces at the metro/non-metro level.

The QLFS uses a rotating sample in which the 3,080 selected PSUs are divided into four rotation groups, from which dwellings are selected. Each sampled dwelling remains in the sample for four consecutive quarters. Rotation is phased so that, at the end of each quarter, one quarter of the sample dwellings are rotated out of the sample and replaced by new dwellings.

In calculating the number and proportion of children living in households with no employed members, *Children Count – Abantwana Babalulekile* in 2009 shifted from using the GHS to using the LFS and the newer QLFS. The labour force-oriented surveys provide a more accurate depiction of employment in South Africa than the GHS, and the derived results are comparable with other analyses such as the government's Development Indicators.

Calculations for each year were backdated to 2003 using the LFS September data for the years 2003 to 2007 and the QLFS 3rd quarter data for 2008. Calculations for 2009 and 2010 are similarly based on 3rd quarter data. Due to the change in survey used, the rates of children living in unemployed households provided in 2009 and later editions are not directly comparable to rates in previous calculations of the *Children Count – Abantwana Babalulekile* project (which drew on data from the GHS).

For more information about the design and limitations of the Labour Force Surveys, see the discussion on the GHS above, and the metadata for the LFS and QLFS, available online.

SOCPEN database[3]

Information on social grants is derived from the Social Pensions (SOCPEN) national database maintained by the South African Social Security Agency (SASSA), which was established in 2004 to disburse social grants for the Department of Social Development. Prior to this, the administration of social grants and maintenance of the SOCPEN database was managed directly by the department and its provincial counterparts.

There has never been a published, systematic review of the social grants database, and the limitations in terms of validity or reliability of the data have not been quantified. However, this database is regularly used by the department and other government bodies to monitor

grant take-up, and the computerised system, which records every application and grant payment, minimises the possibility of human error. Take-up data and selected reports are available from the department on request throughout the year. *Children Count – Abantwana Babalulekile* reports the mid-year grant take-up figures for the sake of consistency with the General Household Survey, which is conducted in June/July each year.

ASSA2008 AIDS and Demographic model[4]

The ASSA2008 suite of demographic models produces time series data on population and HIV-related indicators nationally and by province, population group, sex and age. The models use empirical evidence from surveys and administrative datasets as well as a series of demographic, epidemiological and behavioural assumptions as input. The underlying parameters and assumptions are well accepted and thus the models have been regarded as the "gold standard" in HIV and AIDS, mortality and population projections in South Africa.

ASSA2008, released in March 2010, is the most recent version of the model. It is similar to the previous version, ASSA2003, but has been recalibrated and incorporates some important changes to the underlying assumptions. These include taking into account the slower than anticipated pace of roll out of the Prevention of Mother-To-Child Transmission (PMTCT) programme, allowing for separate antiretroviral treatment (ART) roll-out rates for men, women and children and for higher rates of retention on ART, changes in the way condom usage is modelled and adjusting HIV-survival rates, leading to a longer mean survival time, and even allowing for some infected children to reach adulthood. These changes address some of the limitations of ASSA-2003, amongst which were a tendency to under-estimate paediatric HIV prevalence and survival.[5] The model "represents the triangulation of data from the population census, antenatal survey and registered deaths by some of the country's top actuaries, demographers and epidemiologists".[6]

The ASSA2008 model is available in both a *lite* and *full* version. The *Children Count – Abantwana Babalulekile a*nalyses have been derived from the full version.

References

1 Statistics South Africa (2003 – 2011) *General Household Survey Metadata 2002 – 2010*. Pretoria: Stats SA. Available: http://interactive.statssa.gov.za:8282/webview/.
2 Statistics South Africa (2003 – 2007) *Labour Force Survey Metadata 2003 – 2007*. Pretoria: Stats SA. Available: http://interactive.statssa.gov.za:8282/webview/;
Statistics South Africa (2008 – 2009) *Quarterly Labour Force Survey Metadata 2008 – 2010*. Pretoria: Stats SA. Available: http://interactive.statssa.gov.za:8282/webview/.
3 South African Social Security Agency (2004 – 20011) SOCPEN social grants data. Special request. Pretoria: SASSA.
4 Actuarial Society of South Africa (2010) *ASSA2008 AIDS and Demographic Models*. Available: http://aids.actuarialsociety.org.za/.
5 Shisana O, Rehle T, Simbayi LC, Parker W, Zuma K, Bhana A, Connolly C, Jooste S & Pillay V (eds) (2005) *South African National HIV Prevalence, HIV Incidence, Behaviours and Communication Survey, 2005*. Cape Town: HSRC Press.
6 Dorrington R, Johnson L & Budlender D (2010) *ASSA2008 AIDS and Demographic Models: User Guide*. Cape Town: Actuarial Society of South Africa.