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 approximately 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 2011 GHS 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 10-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 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 10 surveys from 2002 to 2011 suggest that some 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.
- A comparison of the 2011 GHS to ASSA2008 (projected to 2011) suggests that the GHS weights produce an under-estimation of 2% for children below two years and an over-estimation in the region of 7% for children aged 14 – 17 years. For the middle age groups the difference in the estimates is less than 1%. This pattern holds for both boys and girls. The under-estimation is particularly pronounced for babies under a year, at 8%. The male-to-female ratios for all children under 17 are similar across the two sources: 1.00 in ASSA, and 1.01 in the GHS.

The apparent discrepancies in the 10 years of data may slightly affect the accuracy of the Children Count estimates. From 2005 to 2008, consistently distorted male-to-female ratios mean that the total estimates for certain characteristics would be somewhat slanted toward the male pattern. This effect is reduced from 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.

Reporting error

Error may be present due to the methodology used, i.e. 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).

National Income Dynamics Study:[2]

NIDS is the first national panel survey to be conducted in South Africa. The baseline survey, or first "wave" of data collection, was undertaken in 2008, with subsequent waves planned at intervals of two years. In the first wave, data were obtained for every member of each sampled household, and these individuals became the permanent sample members or panel – even if they were children or babies. Subsequent waves endeavour to return not only to the original households, but also to each original household member, even if members have moved out of the household. The advantage of a panel survey is that it enables longitudinal analysis of the variables or outcomes under study, while effectively controlling for variation in individual characteristics.

Wave 1 data collection began in February 2008, and involved 7,305 households and 28, 255 individuals. The primary sampling units were selected from Stats SA's master sample. During the survey, data collected included household demographics, income and expenditure patterns, living conditions, and anthropometric measurements among other indicators.

The anthropometric measurements provide the data for the nutrition analyses in Children Count. To obtain the measurements, fieldworkers recorded two height and two weight measurements for each child, and a third one if the one and two sets of measurements were more than one centimetre or one kilogram apart respectively. An average of the first two measurements was in each case taken for the purposes of Z-scores derivation while the third measure was used for Z-scores derivation if the first two were more than centimetre or one kilogram apart in the height and weight measurements respectively. The weights and heights collected during the study were converted to Z-scores based on the 2006 World Health Organisation's international child growth standards for children aged up to five years.[3] In the case of children older than five years, the WHO growth standards for school-going and adolescent children were used.

Data on height-for-age and weight-for-age covers children aged up to 10 years, while height-for-weight only covers children aged up to five years. In the process of derivation, absolute Z-scores for height-for-age and weight-for-age greater than 6 were treated as biologically implausible and excluded from further analysis.[4] Likewise, absolute Z-scores for weight-for-height of greater than 5 were also found to be implausible and excluded. While it is a nationally representative survey, further disaggregation is limited due to the small sample size used.

SOCPEN database:[5]

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 provides grant take-up figures at the end of March.

References

1   Statistics South Africa (2003 – 2012) *General Household Survey Metadata 2002 – 2011*. Pretoria: Stats SA. Available: http://interactive.statssa.gov.za:8282/webview/.
2   Leibbrandt M, Woolard I & de Villers L (2009) *Methodology: Report on NIDS Wave 1. Technical Paper No.1*. Cape Town: Southern African Labour & Development Research Unit, UCT. Available: www.nids.uct.ac.za/home.
3   Ardington C & Case A (2009) *Health: Analysis of the NIDS Wave 1 Dataset*. Discussion paper no. 2. Cape Town: Southern Africa Labour Development Research Unit, UCT.
4   See no 3 above.
5   South African Social Security Agency (2004 – 20011) SOCPEN social grants data. Pretoria: SASSA.