# 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, 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 multi-stage stratified sampling and probability proportional to size principles. The resulting estimates should be 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 current master sample was used for the first time in 2004, meaning that, for longitudinal analysis, 2002 and 2003 may not be easily comparable with later years as they are based on a different sampling frame. From 2006, the sample was stratified first by province and then by district council. Prior to 2006, the sample was stratified by province and then by urban and rural area. The change in stratification could affect the interpretation of results generated by these surveys when they are compared over time.

*Provincial boundary changes*

Provincial boundary changes occurred between 2002 and 2007, and slightly affect the provincial populations. Comparisons on provincial level should therefore be treated with some caution. 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 Statistics South Africa 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. For this project, GHS data were compared with estimates from the Statistics South Africa's mid-year estimates, as well as the Actuarial Society of South Africa's ASSA2003 AIDS and Demographic model.

Analyses of the six surveys from 2002 to 2007 suggest that over- and under-estimation may have occurred in the weighting process:

- When comparing the weighted 2002 data with the ASSA2003 AIDS and Demographic model estimates, it seems that the number of children aged 0 – 9 years was under-estimated in the GHS, while the number of children aged 10 – 19 was over-estimated. The pattern is consistent for both sexes. The number of very young males aged 0 – 4 years appears to be under-estimated by 15%. Girls in this age group have been under-estimated by 15.8%. Males in the 10 – 14-year age group appear to be over-estimated by 5.7%.
- Similarly in 2003, there was considerable under-estimation of the youngest age group (0 – 9 years) and over-estimation of the older age group (10 – 19 years). The pattern is consistent for both sexes. The results also show that the over-estimation of males (9%) in the 10 – 19-year age group is more than double the over-estimation for females in this age range (3.8%).
- In the 2004 results, it seems that the number of children aged 7 – 12 years was over-estimated by 6%, as well as the number of persons aged 13 – 22 years. The number of very young children appeared to have been under-estimated. The patterns of over- and under-estimation appear to differ across population groups. For example, the number of White children appears to be over-estimated by 14%, while the number of Coloured persons within the 13 – 22-year age group appears to be 9% too low.
- In 2005, the GHS weights seem to have produced an over-estimate of the number of males within each five-year age group. The extent of the over-estimation is particularly severe for the 10 – 14-year age group. In contrast, the weights produce an under-estimate of the number of girls – the error seems greatest in respect of the younger age groups. These patterns result in male-to-female ratios of 1.06, 1.13, 1.10 and 1.09 respectively for the four age groups covering children (ie 0 – 4, 5 – 9, 10 – 14 and 15 – 19 years).
- The 2006 weighting process yielded the same results as in 2005. The one exception is that the under-estimation of females is greatest in the 5 – 9 and 15 – 19-year age groups. This results in male-to-female ratios of 1.03, 1.10, 1.11 and 1.12 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 3 – 5% while the over-estimation is in the range of 1 – 7%. This results in male-to-female ratios of 1.07, 1.06, 1.08 and 1.08 respectively for the four age groups covering children.
- Overall, assuming the ASSA2003 Aids and Demographic model to be the 'gold standard', it appears that the GHS2008 over-estimates both male and female populations under the age of 19 years, except for 0 – 4-year-old females. The extent of over-estimation for boys is in the range 0 – 7%. It is particularly severe for boys aged 10 – 14 years. Over-estimation is in the range of 2 – 5% for girls aged five years and above. For girls aged 0 – 4 years, the ASSA2003 model suggests that these may have been under-estimated by about 1%. The GHS2008 suggests a sex ratio of 1.03 for children aged 0 – 4 years, which is higher than that of the ASSA model and Statistics South Africa's mid-year estimates.

The apparent discrepancies in the seven years of data may slightly affect the accuracy of the *Children Count – Abantwana Babalulekile* estimates. Since 2005 the male and female patterns vary in respect of a particular characteristic, which means that the total estimate for this characteristic will be somewhat slanted toward the male pattern. 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*

Statistics South Africa 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 respon-

dents 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).

**SOCPEN database**[2] Information on social grants is derived from SOCPEN, a national database maintained by the South African Social Security Agency (SASSA), which was established by the government in 2004 to implement the disbursement of social grants for the Department of Social Development. Prior to this, SOCPEN was managed directly by the department. There has never been a published, systematic review of the social grants database, and the extent of the limitations of validity or reliability of the data has 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 GHS survey, which is conducted in June/July each year.

**ASSA2003 AIDS and Demographic models**[3] The ASSA2003 suite of demographic models give time series data on population and HIV-related indicators by province, population group, sex, age and nationally. The models use empirical evidence as well as a series of assumptions as input. The underlying assumptions are well accepted and thus the models have been regarded as the 'gold standard' in HIV/AIDS projections in South Africa. These models give an indication of the proportion of adults and children affected by HIV/AIDS.

Recently, there has been some uncertainty surrounding the models. The ASSA2003 model, which is used to produce the estimates of the annual numbers of new paediatric HIV infections, tends to under-estimate quite substantially the HIV prevalence that has been measured in surveys of older children.[4] This suggests that the annual numbers of new paediatric HIV infections could be under-estimated. For this reason, the indicator 'HIV prevalence among children' has been discontinued in *Children Count – Abantwana Babalulekile*.

There is also substantial uncertainty around the ASSA2003 estimates of the annual numbers of adults progressing to AIDS in each province (the denominator in the calculation of antiretroviral treatment coverage). Caution is therefore required when analysing the relative levels of antiretroviral coverage in the different provinces.

A further limitation, relevant to antiretroviral treatment, is that the ASSA2003 model estimates the number of new AIDS cases rather than the number of individuals who are newly eligible for antiretroviral treatment. The latter includes individuals whose CD4+ counts have dropped below the threshold of 200/μ/, while the former does not. This is likely to imply some under-estimation of treatment need.

In the ASSA2003 model, antiretroviral treatment is assumed to be started at the time of the first AIDS-defining illness, and the calculation of the number of new adult AIDS cases in a particular period is therefore unaffected by the level of antiretroviral provision. Since the ASSA2003 model estimates of annual numbers of new AIDS cases are published over intervals from mid-year to mid-year, the rates of adult antiretroviral coverage are calculated for the same periods.

The ASSA2003 estimates were updated to take into account:

- revised estimates of the proportion of pregnant women who receive HIV counselling and testing (as presented in the section on access to prevention of mother-to-child transmission);
- revised estimates of the proportion of women testing positive who receive nevirapine (this has been set at 75%);
- allowance for the greater effectiveness of the combined AZT and nevirapine regimen that has been introduced in the Western Cape since 2004; and
- revised estimates of the proportion of women who practise exclusive formula feeding.[5]

The model has recently been recalibrated, using more recent data, including the Community Survey 2007. It will be relaunched as ASSA2008 once technical details have been finalised.

**National Comprehensive HIV and AIDS Plan Statistics**[6] This Department of Health report contains the number of adults and children starting antiretroviral treatment in a particular year. The reliability of these data is questionable. For some provinces, like the Northern Cape, the cumulative number of children on antiretrovirals dropped from one year to the other, suggesting data quality problems.

**District Health Barometer**[7] This report by the Health Systems Trust contains data on pregnant women who receive voluntary counselling and testing for HIV. The data show erratic trends in provision of nevirapine to pregnant women and their babies, which may reflect changes in record-keeping rather than quality of service. The data collected from all public health facilities are subject to greater uncertainty and should be treated with caution. There is also provincial variation in the quality of the data. Where provinces produced implausible figures, fields have been left empty. Immunisation coverage is derived from clinic records and reflects the proportion of all children under one-year-old in a target area who complete a primary course of immunisation. Notes on data quality in the *Barometer* suggest some errors in the data from specific hospitals and districts. Some of these data issues are resolved, for instance by removing outliers. Problems with missing denominators seem to have been resolved in 2008.

**National HIV and Syphilis Antenatal Sero-Prevalence Survey in South Africa**[8] South Africa's antenatal clinic data are among the best in Africa. In most other African countries, HIV-prevalence levels are reported in individual clinics or districts, and there is no attempt to draw a nationally representative sample of clinics from which national antenatal clinic prevalence rates can be calculated. This Department of Health survey follows a stratified cluster sampling methodology, with clinics being sampled on a probability proportional-to-size basis. The overall sample sizes are very large, at around 30,000, making this HIV-prevalence dataset one of the largest in the world.

The survey is conducted among pregnant women who attend public health antenatal clinic services during pregnancy. It does not include pregnant women who attend private health facilities, or women who deliver at public health facilities without having made a booking visit. Women seeking antenatal care in the private health sector have a relatively low prevalence of HIV,[9] and thus the surveys over-estimate HIV prevalence in pregnant women generally. It would also be expected that there would be differences in sexual behaviour between pregnant women and non-pregnant women, and the levels of HIV prevalence observed in the antenatal clinic surveys should therefore not be seen as representative of those in the general female population. After controlling for age differences, HIV prevalence in pregnant women tends to be substantially higher than that in women in the general population.[10]

It should also be noted that – in accordance with UNAIDS guidelines[11] – women are tested using a single ELISA antibody test, and there is no confirmatory testing of positive specimens. This may bias the results slightly, as the test can produce false positive results in a small proportion of HIV-negative women. Although this bias is generally thought to be of minimal significance when the population prevalence exceeds 10%, recent South African studies have suggested that the false positive rate could be around 2%.[12] This would imply over-estimation of the true HIV prevalence in pregnant women by about 2%.

**South African Demographic and Health Survey 2003**[13] Two nationally representative South African Demographic and Health Surveys (DHS) have been conducted to date. These cover the population living in private households. The first was conducted in 1998, and the second in 2003. The main survey targets women aged 15 and 49 years. The 2003 survey introduced questions to men on sexual behaviour.

Both surveys use two-stage nationally representative probability samples drawn from Census enumeration areas. The sample is first stratified by the country's nine provinces, and then by urban and non-urban areas. The final sample yielded approximately 10,000 households for 2003.

There was a marked decline in the response rate to the survey. The overall response rate for the women's questionnaire was 75% in 2003, far lower than the 92% in 1998.

The DHS 2003 report suggests an over-representation of urban areas and of the African population group, and an under-representation of Whites and Indian females. It also highlights problems with age misreporting.

Key demographic and health indicators from the DHS 2003 have data quality problems which may be the result of poor fieldwork. These include child mortality, fertility and hypertension prevalence estimates. These indicators are either inconsistent with other data sources or difficult to interpret. Findings that are not sufficiently robust for decision-making are indicated in the report.

The findings on teenage pregnancy rates, sexual behaviour and contraception use must be interpreted carefully. Some of these indicators are affected by the low number of births reported, and by poor data from KwaZulu-Natal. The results are also influenced to some extent by the over-representation of urban areas and Africans.

**National Food Consumption Survey – Fortification Baseline (NFCS-FB) 2005**[14] This study is a cross-sectional survey of a nationally representative sample of children aged 1 – 9 years in South Africa. The survey population consisted of all the children aged 1 – 9 years (12 – 108 months) and women of reproductive age living in the same households in South Africa. This initial sample was adapted by means of 25% over-sampling to accommodate for children and women who would not be home at the time of the survey. A total of 226 enumerator areas (EAs) were included in the survey, 107 of which were urban–formal, 23 urban– informal, 15 rural–formal and 81 tribal areas. All qualifying EAs were selected with a known probability. A qualifying household for inclusion in the survey was defined as any household with at least one child aged between 1 – 9 years and at least one woman of reproductive age living in it.

Validated questionnaires were administered by trained fieldworkers and blood and urine samples were taken from the respondents of each household to assess micronutrient status. Samples of tap water and maize were collected from each household and tested for iodine and vitamin A respectively, the latter at the household level. All questionnaires were translated in the country's official languages for use as appropriate. Quality assurance measures were employed throughout the survey.

For children younger than three years, height was determined by means of a measuring board. The average of two readings was used. If the two readings varied by more than 0.5 cm, the procedure was repeated. For children three years of age and older, height was determined by means of a stadiometer. The average of two readings was used. If the two readings varied by more than 0.5 cm, the procedure was repeated.

Weight was determined for all children using pre-calibrated electronic scales. The average of two readings was used. The procedure was repeated once. The two readings could not vary by more than 100g; if so, the scale was checked for accuracy and the procedure repeated.

Vitamin A status was classified according to the World Health Organisation's criteria. Status was determined on the basis of the serum vitamin A concentration present in the blood drawn from children in the sample.

## References


1 Statistics South Africa (2003 – 2009) *General Household Survey Metadata 2002 – 2008*. Pretoria: StatsSA. Available: www.statssa.gov.za.
2 South African Social Security Agency (2004 – 2009) SOCPEN social grants data. Pretoria: SASSA.
3 Dorrington RE, Johnson LF, Bradshaw D & Daniel T (2006) *The demographic impact of HIV/AIDS in South Africa. National and provincial indicators for 2006*. Centre for Actuarial Research (UCT), Medical Research Council and Actuarial Society of South Africa.
4 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.
5 Doherty T, Besser M, Donohue S, Kamoga N, Stoops N, Williamson L & Visser R (2003) *An evaluation of the prevention of mother-to-child transmission (PMTCT) of HIV initiative in South Africa: Lessons and key recommendations*. Durban: Health Systems Trust.
6 Department of Health (2008) National comprehensive HIV and AIDS plan statistics. [Unpublished].
7 Health Systems Trust (2009) *District Health Barometer*. Available: www.hst.org.za.
8 Department of Health (2001; 2009) *National HIV and Syphilis Prevalence Survey 2000; National HIV and Syphilis Prevalence Survey 2008*. Pretoria: DoH.
9 Wilkinson D (1999) HIV infection among pregnant women in the South African private medical sector. *AIDS*, 12(13): 1783.
10 See no. 4 above; Connolly C, Shisana O, Colvin M & Stoker D (2004) Epidemiology of HIV in South Africa – Results of a national, community-based survey. *South African Medical Journal*, 94(9): 776-781.
11 World Health Organisation (2002) *Scaling up antiretroviral therapy in resource-limited settings: Guidelines for a public health approach*. Geneva: WHO.
12 Amirfar S, Hollenberg JP & Abdool Karim SS (2006) Modeling the impact of a partially effective HIV vaccine on HIV infection and death among women and infants in South Africa. *Journal of Acquired Immune Deficiency Syndromes*, 43(2): 219-225.
13 Jackson DJ, Chopra M, Doherty TM, Colvin M, Levin J, Willumsen J, Goga A & Moodley P (2007) Operational effectiveness and 36 week HIV-free survival in the South African programme to prevent mother-to-child transmission of HIV-1. *AIDS*, 21(4): 509-516.
14 Labadarios D (ed) (2007) *The National Food Consumption Survey – Fortification Baseline (NFCS-FB): The knowledge, attitude, behaviour and procurement regarding fortified foods, a measure of hunger and the anthropometric and selected micronutrient status of children aged 1 – 9 years and women of child bearing age: South Africa, 2005*. Pretoria: Directorate: Nutrition, Department of Health.