Technical notes on the data sources

General Household Survey: 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, 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. 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 (Statistics South Africa 2008).

Weights
Person and household weights are provided by Statistics South Africa and are applied in the analyses to give estimates at the provincial and national levels.

GHS data were compared with estimates from the Statistics South Africa's mid-year population 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%.
- 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.

The apparent discrepancies may slightly affect the accuracy of the *Children Count — Abantwana Babalulekile* data. For example, where the male and female patterns vary in respect of a particular characteristic, 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 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).

SOCPEN database: 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.

Education statistics in South Africa at a glance: Learner-to-educator ratios and the Gender Parity Index are based on the Department of Education's annual SNAP survey, conducted across the country on the tenth day of the school year. As this survey is conducted annually, data should be available on a yearly basis, but are usually released two years later.

The data collection and processing of this survey have been known to be problematic, and the accuracy and reliability of this data is therefore questionable. The Education Department has previously noted this as a problem, and there have been efforts to improve quality controls in recent years. The department signed the Protocol for Inter-Governmental Cooperation with Statistics South Africa, which means that data must comply with quality standards in order to be accredited as official national data. Stats SA's 'Statistics Quality Assurance Framework' provides data quality guidelines and monitors the quality of the statistics produced in the country. This may help to improve data quality.

The department launched a new administrative data system, the National Learner Unit Record Information and Tracking System (LURITS) in September 2008. The system should provide individual records of enrolment from 2010, and may eventually enable the monitoring of learners' attendance and progress, even if they change schools.

School Register of Needs and National Education Infrastructure Management System: Data for school water and sanitation draw on the 2000 School Register of Needs and the 2006 National Education Infrastructure Management System.

The 2000 School Register of Needs survey collected information from 27,148 public and independent schools — covering more schools than the previous (1996) survey. School principals completed the survey forms themselves, and this may have influenced the objectivity of reporting. Provincial departments were required to verify the data provided by schools in their province. The survey was conducted in eight of the nine provinces, while Mpumalanga conducted its own survey. This may have influenced the national results, although there were attempts to control for variation.

The National Education Infrastructure Management System (NEIMS, 2006) collected information from 30,117 education sites comprising public schools, public early childhood education centres, public ELSEN centres (special needs), public Adult Basic Education Training centres and educational offices of the Department of Education. Of these education sites, 25,145 were public schools. Independent schools were excluded from the assessments, or specified separately.

Information on the data collection processes is not readily available; therefore the quality of the data cannot be easily ascertained.

The Department of Education (2007) provides the following caveats:
- The master list of education sites is regularly improved and may change as new schools are established.
- These statistics should not be interpreted to mean that infrastructure is either at the appropriate level of service or in an acceptable condition.
- There have been differences in the definition of various parameters from previous School Register of Needs assessments. These include level of water supply and type of sanitation.

Data for different years are not directly comparable:
- There are known errors and omissions in the School Register of Needs data. The numbers do not add up to the total number of schools. It is not clear why there are more schools counted in 2000 and in 1996 than in 2006.
- The 2006 data (NEIMS) reflect public (government) schools only, while the 2000 and 1996 data (School Register of Needs) included both public and independent schools.
- In the 2006 NEIMS report, the category "pit latrines/enviroloo" was divided into two categories, namely "[ordinary] pit latrines" and "VIP & enviroloo toilets". These figures cannot be compared with those from the Schools Register of Needs, where all three types were collapsed into a single category.
- The School Register of Needs 1996 and 2000 reported "water on site", while the 2006 NEIMS reported "water on or near site". It is not possible to determine the extent to which an apparent increase in schools with access to water is the result of different (broader) question formulation or the expansion of water service provisioning to schools.
- In the 2006 NEIMS report only, the category "water on or near site" was divided into two sub-categories, namely "schools depending on boreholes on site or rainwater harvesting systems" and "schools served by the municipality". In the *Children Count — Abantwana Babalulekile* indicator the two categories have been collapsed to indicate "schools with water on site or near site" for comparison with previous years.

South African Burden of Disease Study: This 2000 study by the Medical Research Council makes use of vital registration data (number of official births and deaths) but adjusts for under-registration, as large numbers of births and deaths of younger children in particular are unreported. A modelling approach, developed by the Actuarial Society of South Africa (ASSA) was then used to estimate the total number of deaths, since vital statistics are incomplete. The ASSA2000 model was

used to determine overall mortality, the population size, and the number of deaths due to HIV/AIDS for each province.

The basic mortality assumptions for children were as follows: "Child mortality estimates from the 1996 Census and the 1998 *Demographic Health Survey* (SADHS) both show a reversal of the downward trend, although there are differences in the estimated levels (Nannan et al, 2000). Adjustments are made to both sets of estimates due to differences and inherent biases in the different methodologies. A small upward adjustment is made to the DHS and a downward adjustment to the Census data which appear too high due to the inclusion of stillbirths incorrectly classified as live births who have died (Moultrie and Timæus, 2002)". The ASSA-modelled estimates are made available on a yearly basis.

ASSA2003 AIDS and Demographic models: 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 (Dorrington, Bradshaw, Johnson & Budlender 2004). 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 (Shisana *et al* 2005). 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/µl, 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 have been 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 (Doherty *et al* 2003).

National Comprehensive HIV and AIDS Plan Statistics: This report by the Department of Health 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 Northern Cape, the cumulative number of children on antiretrovirals dropped from one year to the other, suggesting data quality problems.

District Health Barometer: This report by 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. Leigh Johnson of the Centre for Actuarial Research at UCT suggests that it 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. Some provinces produced implausible figures; thus these fields have been left empty.

National HIV and Syphilis Antenatal Sero-Prevalence Survey in South Africa: 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 (PPS) 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 (Wilkinson 1999), 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 (Shisana *et al* 2005; Connolly *et al*, 2004).

It should also be noted that — in accordance with UNAIDS guidelines (UNAIDS/WHO 2001) — 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% (Jackson *et al* 2007; Amirfar *et al* 2006; Johnson *et al* 2007). This would imply over-estimation of the true HIV prevalence in pregnant women by about 2%.

Sources

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.

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.

Health Systems Trust (2009) *District Health Barometer*. Available: www.hst.org.za.

Department of Education (2007) *National Educational Infrastructure Management System* (NEIMS). Pretoria: DoE.

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.

Dorrington RE, Bradshaw D, Johnson L & Budlender D (2004) *The Demographic Impact of HIV/AIDS in South Africa. National Indicators for 2004*. Cape Town: Centre for Actuarial Research, Medical Research Council & Actuarial Society of South Africa.

Dorrington R & Kramer S (2004) *The 2004 mid-year estimates: Method, reliability and implications*. Paper presented at a Centre for Actuarial Research seminar, UCT.

Dorrington R & Bourne D (2008) Has HIV prevalence peaked in South Africa? Can the report on the latest antenatal survey be trusted to answer this question? *South African Medical Journal*, 98 (10): 754-755.

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.

Johnson LF, Dorrington RE & Matthews AP (2007) An investigation into the extent of uncertainty surrounding estimates of the impact of HIV/AIDS in South Africa. *South African Journal of Science*, 103: 135-140.

Moultrie T & Timæus I (2002) *Trends in South African fertility between 1970 and 1998*. Cape Town: Medical Research Council, Burden of Disease Unit.

Nannan N, Bradshaw D, Timæus I & Dorrington R (2000) *The impact of HIV/AIDS on infant and child mortality in South Africa*. XIIIth International AIDS Conference, Durban, South Africa, 8 – 13 July 2000. [Abstract of poster MoPeD2507]

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.

Statistics South Africa (2007) *Mortality and causes of death in South Africa, 2005. Findings from death notification*. Pretoria: StatSA.

Statistics South Africa (2003 – 2008) Metadata for *General Household Survey 2002; General Household Survey 2003; General Household Survey 2004; General Household Survey 2005; General Household Survey 2006; General Household Survey 2007*. Pretoria: StatsSA. Available: www.statssa.gov.za.

South African Social Security Agency (2004 – 2008) SOCPEN social grants data. Pretoria: SASSA.

Wilkinson D (1999) HIV infection among pregnant women in the South African private medical sector. *AIDS*, 13 (13): 1783.