Technical notes on the data sources

General Household Survey: The GHS is an annual survey conducted by the national statistics body, Statistics South Africa (www.statssa.gov.za). The sample used is based on the enumeration areas established during the Census demarcation phase and therefore covers all parts of the country. The sample of 30,000 dwelling units ensures as much representivity as possible by stratifying by province, and then by urban and rural area. The resulting estimates should be representative of the total population of South Africa. A weighting process is also applied to improve the representivity of the estimates. These weighted results are used for the *Children Count – Abantwana Babalulekile Project*.

However, over- and under-estimation appears to have occurred in the weighting process:

- The 2002 weighting process appears to have under-estimated the youngest age group (0 – 9 years), and over-estimated the older age group (10 – 19 years) relative to the *ASSA2003 Aids and Demographic* estimates. The pattern is consistent for both sexes. The number of very young males aged 0 – 4 years appears to be under-estimated by 15%. Similarly, girls in this age group have been severely under-estimated (15.8%). Males in the 10 – 14-year age group appear to be over-estimated by 5.7%.
- Similarly in 2003, considerable under-estimation at the youngest age group (0 – 9 years) and over-estimation at the older age group (10 – 19 years) have occurred. 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 extent of 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.
- 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 apparent discrepancies in the five years of data will affect the accuracy of the *Children Count – Abantwana Babalulekile* data. For 2005 and 2006 where, for example, the male and female patterns in respect of a particular characteristic vary, the total estimate for this characteristic will be somewhat slanted towards 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.

Further 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 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).

In general, the GHS questionnaires for the five-year period are very similar in respect of the questions used for the *Children Count – Abantwana Babalulekile* indicators. Comparison of results of the 2002 – 2006 surveys, including comparisons of the extent to which answers are "unspecified", does not suggest any noticeable impact on quality.

The surveys do not cover other collective living-quarters such as students' hostels, old-age homes, hospitals, prisons and military barracks. It does cover workers' hostels. The exclusions should not have a noticeable impact on the findings in respect of children.

Confidence intervals will be available for the major categories of data for the five years of data presented on the project's website. Those indicated in the data tables refer to intervals wider than 5%.

***SOCPEN database*, Department of Social Development:** There has never been a published, systematic review of the SOCPEN database, and the extent of the limitations of validity or reliability of the data has not been quantified. However, it is regularly used by the department and other government bodies to monitor grant take-up. This administrative dataset is constantly updated by Department of Social Development employees when entering application and payment data. Take-up data and selected reports are available from the department on request throughout the year. Grants data will be updated regularly for the *Children Count – Abantwana Babalulekile Project*.

***Education statistics in South Africa at a glance*, Department of Education:** This data is based on the department's annual survey and SNAP ('snap-shot') survey, taken on the tenth day of the school year. The data capturing and processing of this survey are known to be problematic and erroneous, although the data quality seems to be improving. The accuracy and reliability of this data is therefore questionable.

As this survey is conducted annually, data should be available on a yearly basis. However, data processing systems differ

across the provinces, and some are more efficient than others. The department's current information management system, known as the Education Management Information System (EMIS), is presently under review.

Causes of death in South Africa 1997 – 2001, Advance release of recorded causes of death; Mortality and causes of death in South Africa, 2003 and 2004: Findings from death notification; and Mortality and causes of death in South Africa, 2005: Findings from death notification: This data is obtained from Statistics South Africa's regular statistical releases on mortality and causes of death in South Africa. The number of deaths reflected in the data excludes stillbirths. The data captured in Statistics South Africa's statistical releases are obtained from death notification forms from the Department of Home Affairs. There are a number of factors related to these forms that limit the accuracy and completeness of the data. For example, the data obtained on these forms are subject to content errors and omissions. Under-registration of deaths occurs, particularly in rural areas and among children. Furthermore, the causes of death may be misreported on the form. Statistics South Africa (2007) notes in particular that the codes used to classify deaths for children younger than one year should be treated with caution, as they do not take into account the exact age at death for infants.

ASSA2003 AIDS and Demographic Model: Currently the only available data on HIV-related indicators focusing on all children are estimates based on modelling. The underlying assumptions of the model, however, are well accepted nationally and these are thus the best estimates available at present.

Estimates are obtained by using mathematical models. These models give an indication of the proportion of adults and children affected by HIV/AIDS. The demographic model is based on a wide range of available empirical evidence, for example, regular survey data and vital statistics, such as the antenatal clinic survey results and number of deaths from the population register (Dorrington, Bradshaw, Johnson & Budlender 2004). Data and modelled results are available at www.assa.org.za.

***National HIV and Syphilis Antenatal Sero-Prevalence Survey in South Africa*, Department of Health:** This study was conducted as an anonymous survey among pregnant women who attended public health antenatal clinic services for the first time during pregnancy. Sentinel sites were selected based on the Probability Proportional to Size (PPS) sampling method. A stratified proportional sample was drawn and the sample size was proportionally allocated to each facility. By 2006, the sample size was 36,000. These studies were limited by several factors. As the study is conducted in public health facilities, the sample is not necessarily representative of the demographic and socio-economic profile of the country.

Sources

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, South African 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, University of Cape Town.

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