AN EVALUATION OF AGE-SEX DISTRIBUTIONS OF SOUTH AFRICA'S POPULATION WITHIN THE CONTEXT OF HIV/AIDS

By

Eric O. Udjo

Research Director

Epidemiology and Demography Unit

Social Aspect of HIV/AIDS and Health

Human Sciences Research Council
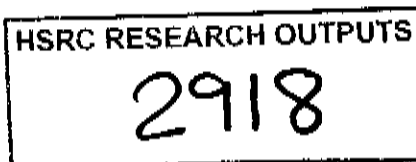
Private Bag X41

Pretoria 0001.

Tel: (27) (12) 302 2603

Fax: (27) (12) 302 2601

e-mail: udjoeric@hotmail.com

AN EVALUATION OF AGE-SEX DISTRIBUTIONS OF SOUTH AFRICA'S POPULATION WITHIN THE CONTEXT OF HIV/AIDS

ABSTRACT

Knowledge about the age-sex distribution of a population is valuable for planning and decision-making as it is a determinant of the future development of the population. Various aspects of the age-sex distributions of the population of South Africa estimated by Statistics South Africa remain controversial. For example, at face value, the 1996 census implies that overall sex ratio dropped from 97 in 1970 to 92 in 1996 and 91 in 2001. Also the 1996 census results imply that the proportion of persons aged 0-4 dropped from 15% in 1970 to 11% in 1996 and 10% in 2001. As an attempt to shed more light on these controversies, this study evaluates the age-sex distributions from the 1996 census using a novel approach. The study also develops model age-sex distributions for South Africa as a means of evaluating age-sex distributions from the 2001 census, and future censuses and surveys in the country.

AN EVALUATION OF AGE-SEX DISTRIBUTIONS OF SOUTH AFRICA'S POPULATION WITHIN THE CONTEXT OF HIV/AIDS

INTRODUCTION

The age-sex distribution of a population is a crosscutting variable in planning as it is intrinsically linked to all aspects of the life cycle including childhood, education, marriage, childbearing, entry into the labour market, retirement, ageing, morbidity and mortality. Knowledge about the age-sex distribution of a population is therefore valuable information for planning and decision-making as it is also a determinant of the future development of the population. Planning and policy decision can be impaired due to inaccurate information about the age-sex distribution of the population. For example, in the education sector, incomplete coverage of persons in certain age groups in a census if not corrected for can result in the underestimation of the school age population in population projections and which may result in inadequate allocation of resources in planning for this sector. Likewise, incomplete coverage of persons in certain age groups by sex (e.g. the 0-4 or 20-24) in a census if not corrected for, could lead to erroneous conclusions about the impact of HIV/AIDS which might impact on planning in the health sector. It is therefore essential to appraise age-sex distributions from a census or survey to accurately inform planning and decision-making.

3

Various aspects of the age-sex distributions of the population of South Africa estimated by Statistics South Africa (1999) from the 1996 census remain controversial. (See Dorrington 1999; Sadie 1999; Udjo 1999, Shell 1999; Phillips et al. 1999). Some of the controversial issues include: At face value, the 1996 and 2001 censuses imply that overall sex ratio dropped from 97 in 1970 to 92 in 1996 and 91 in 2001. Furthermore, the results of the censuses imply that the proportion of persons aged 0-4 dropped from 15% in 1970 to 11% in 1996 and 10% in 2001. As an attempt to shed more light on these controversies, this study evaluates the age-sex distributions from the 1996 census using a novel approach that had hitherto not been undertaken. Also, the study develops model age-sex distributions for South Africa as a means of evaluating age-sex distributions from future censuses and surveys in the country.

CONCEPTUAL PROBLEM IN EVALUATING AGE-SEX DISTRIBUTIONS

An age-sex distribution is determined by past fertility, mortality and migration. Perhaps the most disturbing problem in the evaluation of age-sex distributions is migration. International and internal migration data are problematic. Firstly, documented international migration (emigration and immigration) at any point in time may be incomplete because people may not declare their intentions on entry or exit at the designated entry/exit points. Secondly, because national borders are porous, the magnitude of undocumented migrants who may be residing in a country is generally unknown. This poses a problem in the evaluation of age-sex

distribution at the national level, as it is difficult to ascertain whether or not certain observed features in age-sex distributions are due to international migration.

Thirdly, few countries in the world track internal migration through a system of population register. Although internal migration data may be available from surveys and censuses, the accuracy of such data is often in doubt thus limiting age-sex evaluation at the sub-national level.

These problems exist in varying degrees in South Africa. For example with regard to documented emigration, the official figures are unreliable. Estimates of the ratio of South African emigrants (reported in major receiving countries) to South Africa's Home Affairs figures is 1.4 for the period 1970-96 (Bah, 1999); 2.8 for the period 1989-97 (Kaplan et al, 1999); 1.4 – 4.7 (Interim Statistics Council 1996 Census document). Additionally, the scale of undocumented immigration (presumably largely African) is unknown. McDonald et al (1999), Brunk (1996) for example, quote some figures that have been speculated. Crush, (1999; cited in McDonald, 1999) also provides some figures. The figures range from half a million to 8 million. In view of this, it is extremely difficult to assess and control for the impact of migration in the evaluation of age-sex distributions of South Africa's population.

One approach in controlling for migration is to exclude non-South African citizens from the analysis. This approach is however unsatisfactory because of

uncertainty surrounding the accuracy of available figures. With regard to the 1996 census for example, a total of 369,950 (i.e. 1% of the total population) were reported as non-South Africans. This figure is negligible but has been contested by a number of researchers as an underestimate. It has been argued for example that a process of "self naturalization" among undocumented immigrants may have been at play during enumeration in the 1996 census for fear of repatriation (see Interim Statistics Council document).

In view of the uncertainty regarding the volume of migration, migrants were not excluded from the analysis in this study. Hopefully, the volume of net international migration and its impact on the age-sex structure in South Africa are relatively negligible in comparison with the other demographic processes. It, is possible however that this study underplays the impact of net international migration on the age-sex distributions.

DATA

The sources of data for this study are the 1970 and 1996 censuses, and the 1998 October household survey (OHS98).

The 1970 and 1996 censuses

The choice of the 1970 and 1996 censuses for this study was based on the following consideration. The 1970 census was the last census during the apartheid era that canvassed the entire country while that of the 1996 was the first post-apartheid nation-wide census. The censuses between these two periods were fragmentary in the sense that they did not canvass the entire country. While the 1970 census had an estimated undercount of about 2% (see Sadie, 1988), that of 1996 had an estimated undercount of about 11%).

The census questionnaires in both the 1970 and 1996 censuses included the pertinent variables for this study – age, sex and population group (African/Black, coloured, Indian and white). In addition, the 1996 census had information on region of residence – province – as defined in the post-apartheid era, which is different from regional boundaries in the apartheid era. The estimated population in 1970 was 22.1 million (see Sadie, 1988) while it was estimated as 40.6 million in 1996 (see Statistics South Africa, 1999). Because of the controversies surrounding the adjustments of the 1996 census (and which has resurfaced in

recent year), the analysis in this study utilising 1996 censuses was based on the unadjusted figures.

1998 October Household Survey

Statistics South Africa began a programme of October Household Surveys in 1994. Until 1998, they were omnibus surveys. On the basis of the assessment of various aspects of quality in the various surveys by Udjo (1997, 1998 and 2003), the 1998 survey of all the October household surveys is the most appropriate for this study.

The sample size for the 1998 survey was 20,000 households. Though a probability sample, was not self weighting nationally. Analysis of the data was therefore carried out at levels (population group, place of residence) where they were roughly self-weighting. Hirschowitz and Orkin (1996) provide details regarding the design of the October household surveys.

METHODS

The analysis in this study is based on the following premise. In a closed population, the sex ratio in specified age categories derived from the listing of household members should be equal to that derived from mothers residing in the household reporting on the number of child children alive and living with the

8

mother at the time of enumeration in the absence of errors in both sets of report. While the listing of members of households is often obtained from the head of household (or someone acting in that capacity) in a census or survey household questionnaire/schedule, the latter is often obtained from birth histories in a survey questionnaire. Comparison of sex ratios in specified age categories derived from both sets of data can be used to appraise the quality of either or both sets of data.

Careful consideration must be given in the choice of the age categories on which the sex ratios are based. This is because the age-sex reports from the household questionnaire spans the entire age range in the population while the age-sex report on children alive (and living with the mother) from birth histories only covers a maximum age that is about 15 years lower than that of the oldest mother living in the household listed in the household schedule.

In theory, the persons in birth histories are the sons and daughters of mothers listed in the household and if alive and living with the mother, would comprise a substantial component of those listed in the household. On the other hand persons listed in household can be categories broadly into the following: (1) Non-family members (e.g. visitors, domestic workers); (2) Sons and daughters of the household head and/or his/her partner; (3) Other family members (e.g. head of household, wife/husband/partner, father of head of household etc. In general (1) above is numerically small compared to (2) and (3). In developing countries

9

generally, children (or persons under the age of 15) constitute a very substantial part of the population. Thus in any population, a strong driving force in sex ratios in certain age groups is the second component above.

If there is no marked trend in sex ratios (due to trends in sex ratio at birth, differential mortality and migration by sex) in the population, sex ratios in certain age groups derived from a household schedule in a survey/census should be approximately similar to the sex ratios derived from the birth histories in the same survey. Sex ratios usually do not show a marked trend in a short period of time (say five to ten years).

Sources of discrepancy in the sex ratios from both sets of reports and over time, may arise from one or more of the following: (1) differential coverage by sex; (2) magnitude of, and changing pattern of age/sex differential in mortality; (3) differential pattern of age misreporting by sex; (4) changing pattern of migration by sex.

As noted, migration is a complicating factor in the analysis of sex ratios. However, this is not important when assessing quality of two sets of report from the same survey. If however, one is comparing sex ratios from two data sources at different time periods, migration cannot be ignored, and was controlled for in this study. Differential pattern of mortality and age misreporting were also controlled for thus, some of the results presented in this study is an indication of

10

the magnitude of the discrepancy in sex ratios attributable mainly to differential coverage by sex.

Sex ratio computed, as the number of males per 100 females is a crosscutting measure in this study. The effect of differential age misreporting on sex ratios was minimised by computing the ratios for broad age groups (less than 15 years, 15-64 and 65+ age groups). An overall ratio was also computed for persons aged 0-64. The choice 64 as the upper age limit was based on the following consideration. Since no upper age limit was imposed on women to whom the birth histories were administered in the 1998 survey, the children in their 70s at the time of the survey would have been born by mothers in their late 80s at the time of the survey. However, to minimise the effect of random fluctuation due to small numbers in the survey and to make the sex ratio in the birth history comparable to that obtained from the household schedule, an upper age limit of 64 was imposed. (Note that the age band 0-64 covers about 95% of South Africa's estimated population). To minimise the effect of migration at the household level, the computation of sex ratios from the birth histories was restricted to children alive and living with the mother at the time of the survey so that it is comparable to the sex ratio obtained from the household schedule. In the presentation of the results, sex ratio from birth histories relates to children alive and living in same household as the mother at the time of the survey. Note that since censuses (with a few exceptions e.g. in Namibia) do not collect birth

11

histories, sex ratios from birth histories relate only to the 1998 October household survey.

Smoothing of the reported 1970 age-sex distributions was carried out by comparing the logit transformations of the reported cumulated age-sex distribution with an appropriate stable population. The selection of the stable populations was based on the similarity of estimated parameters (including the general fertility rate, proportions under the age of 15 and 45+, life expectancy at birth and the rate of population growth) of the observed population in comparison with those of the stable population (see Hobraft for the construction of the stable populations).

The logit transformations, Y_x and Y_{sx} given by Brass (1971) are denoted

$Y_x = 0.5\log_e((1-y_x)/y_x))$ and
$Y_{sx} = 0.5\log_e((1-y_{sx})/y_{sx}))$

Where y_x is the cumulated proportion under age x in the observed population. The subscript s, denotes the stable population. The Y_x values were plotted against the Y_{sx} values and a straight line fitted to the "best" points using the ordinary least squares method.

The smoothed age distribution, $Y_{x'}$ was computed as

$$Y_{x'} = \alpha + \beta Y_{sx}$$

Where the terms on the right hand define the fitted line. Re-transformation of the fitted values, was obtained using the expression

$$Y_{x't} = 1/(1+e^{2Yx'})$$

Where $Y_{x't}$ is the antilogit of $Y_{x'}$. $Y_{x't}$ was decumulated to obtain the smoothed proportion of persons in age group x. This proportion was then applied to the total number of persons (by sex) adjusted (by Sadie, 1988) for undercount in the census to obtain the smoothed number of person in age group x.

The AIDS impact model (AIM) developed by the Futures Group was used to incorporate the impact of HIV/AIDS in the population estimates. The method is described by Stover (1999). Estimation of the basic parameters required by AIM was based on the following assumptions. First it was assumed that the Nelson Mandela/HSRC (2002) data provide a reasonable measure of HIV sero-prevalence at the national level and hence the Department of Health's antenatal sero-prevalence for the period 1990-2001 calibrated by Rehle and Shisana (2003) provide a reasonable measure of national adult HIV sero-prevalence for the period 1990-2001. Second, it was assumed that the distribution of national adult HIV sero-prevalence by population group in 2002 as measured by Nelson

13

Mandela/HSRC study is similar to the distribution of adult sero-prevalence by population group in previous years. On this basis, the ratio of the adult HIV sero-prevalence of each population to the total adult prevalence in 2002 was used as weight on the prevalence for each year to obtain an estimate of each population group for each year during the period 1990-2002. Lastly, it was assumed that the age specific sero-prevalence among adults in the Nelson Mandela/HSRC data is similar for all population groups and is a reasonable measure of age-specific sero-prevalence in 1990-2002.

Immigration and emigration statistics are available from the Department of Home Affairs through Statistics South Africa for the estimation period (1970-2001). However, the statistics are not given by population group; also, the number of documented emigrants appears to understate the true extent of the phenomenon as already. In view of this, net migration was estimated assuming that: (1) the number of documented immigration is accurate. (2) The percentage distribution of documented immigrants by population group is similar to the percentage distribution of non-South Africans by population group and sex as reported in the 1996 census (though the number may not be accurate). (3) The number of documented emigrants is inaccurate. (4) Prior to 1994,° the percentage distribution of documented emigrants by population group is similar to the percentage distribution of non-South Africans by population group as reported in the 1996 census. Only whites have been emigrating since 1994 (the year of the new political dispensation). (5) The age distribution of documented immigrants

for the period 1996-2001 is a reasonable estimate of the age distribution of net migrants for the estimation period.

On the above basis, the documented volume of immigration during the period 1970-2001 was split by population group and sex. Regarding emigration, a correction factor of 1.4 was used to adjust the number of male female documented number of emigrants for the period 1970-93 while for the period 1994 to 2001, a correction factor of 2.7 was applied (see Bah, 1999; Kaplan et al, 1999). The adjusted figures for the period 1970-93 by sex were then split by population group on the basis of the above assumptions. The adjusted figures by sex for the period 1994 to 2001 however, were attributed to the white population. An average of the male-female age distribution of documented immigrants for the period 1996-2001 was used as the best estimate of net migrants for the entire estimation period.

RESULTS

Internal evaluation of the overall Sex ratio from the 1998 survey

Some researchers (Sadie, 1999; Dorrington, 1999) have argued that the overall sex ratios from the 1996 census are too low and indicative of a general under enumeration of males during the census.

The first panel of Table 1 shows a consistent difference in the overall ratio (i.e. for person aged 0-64) as reported in the household schedule and the birth history in the 1998 October household survey among the four population groups. The ratios are lower in the household schedule as compared with the birth history with Africans showing a very low sex ratio. Since the two sets of reports were obtained in the same survey, the difference cannot be attributed to migration or differential pattern of age-sex mortality but point to differential coverage (which could be in either direction) in either set of data. While the ratios from the birth histories are close to 100 in the four population groups and similarly, among Indians and white in the household schedule but not among Africans and coloureds. This appears to suggest under reporting of males in the household listing especially among the African population during the 1998 survey. An alternative interpretation is over statement of female Africans especially, due to mis-statement of sex.

Further decomposition of the sex ratios of the 0-64 age band revealed that the main source of the discrepancy in the sex ratios derived from the household scheduled compared with those derived from the birth history is in the sex ratio of the 15-64 age group. Among Africans for example, the sex ratio of the 15-64 age group is 82 in the household schedule compared to 97 in the birth history (see third panel of Table 1). The high sex ratio for the 15-64 age group from the birth histories of Indian women is probably due to selective omission of dead females in their reports.

Four of the nine provinces (Eastern Cape, Kwazulu-Natal, Limpopo and Mpumalanga) also exhibit substantial inconsistencies in the sex of ratios of persons aged 0-64 derived from the household schedule compared with those derived from the birth history of the 1998 survey (see Figure 1). The sex ratios from household schedule are very low sex (due to the low sex ratios in the 15-64 age group). In Limpopo province for example, whereas the sex ratio for the 0-64 age group is 81, the corresponding value from the birth history is 104. The inconsistency may not be attributable to migration since the two sets of information were collected from the same unit of enumeration in the same survey. It appears from this that the apparent under-enumeration of males (or over statement of the number of females) in the household listing during the 1998 survey was more marked in Eastern Cape, Kwazulu-Natal, Limpopo and Mpumalanga compared with the other provinces.

In these provinces, there are also marked discrepancies in the sex ratios of the 0-64 age group derived from the birth histories of the 1998 survey compared with the ratios derived from the 1996 census for the same age group, which may not be attributed to migration. If the discrpancies were due to migration, one would expect low sex ratios of the 15-64 age group obtained from the birth histories but this is not case as the birth histories indicate much higher sex ratios in the 15-64 age group compated with those obtained from the 1996 census for the same age

17

group in each of the provinces. (It is unlikely that there could have been systematic reverse migration in two years).

Consistency of reported overall sex ratios overtime as a means of evaluating the sex ratios from the 1996 census

As seen in the first panel of Table 1, the overall sex ratios (0-64 age group) from the 1996 household members list among Africans and coloureds are inconsistent with those from the birth histories of the 1998 survey and the 1970 census especially among Africans. Among Africans, the reported ratio was 98 in the and 93 in the 1970 and 1996 censuses respectively, and 99 from the birth history of the 1998 survey. It is inconceivable that demographic processes (i.e. fertility, mortality and migration) resulted in a sharp decline in the sex ratio from a high of 98 in 1970 to a low of 93 in 1996. The ratio from the birth history of the 1998 survey is consistent with the ratio in 1970. The low ratio from the 1996 census could either be due to under-enumeration of males or over statement of the number of females due to mis-classification of sex during the listing of household members in the 1996 census.

Under-enumeration of male or over enumeration of female Africans and coloureds during the 1996 census appears to have been marked in the working age group compared (see the last two panels of Table 1).

18

The 0-4 age group

Indentation at the base of the population pyramid is a consistent feature of reported age-sex distributions in various household surveys and censuses undertaken by Statistics South Africa. Figures 2 and 3 appear to suggest that the proportion aged 0-4 increased by more than 1% between 1995 and 1996. Furthermore, Figures 4 and 5 show marked differences in the age-sex distributions in two different surveys that were undertaken in the same yeare in 1997. Both pyramids appear to suggest that within the same year in 1997, the proportion aged 0-4 changed by nearly 3%. Comparing Figures 2 and 6, one would expect that the reported cohort 0-4 years old in 1995 (9.6%) would 5 years later in 2000 move to the cohort 5-9 years old and taking account of mortality, the proportion aged 5-9 in 2000 would be less than 9.6% in the absence of large scale immigration in this particular age group (i.e. 5-9 in 2000). In contrast, the reported proportion aged 5-9 in 2000 (11.2%) is higher than the reported proportion aged 0-4 in 1995 (9.6%). These patterns are not consistent with fertility trend in South Africa.

The ratio of the 0-4 and 5-9 and groups can be used to evaluate the accuracy of the reported proportions aged 0-4 in the 1998 survey and the 1996 census. In view of levels and trends in fertility in South Africa (see Udjo, 2003), the ratio of the 0-4 and 5-9 age groups should be greater than unity   among Africans and

coloureds in 1998 and 1996. A ratio of these age groups less than unity is therefore indicative of the minimum extent of underreporting of the 0-4 age group in a particular year (unless if mortality was higher among children 0-4 compared to children of similar age five years before). With regard to white and Indians, one would expect a ratio less than but close to unity in view of their low levels of fertility over the years.

As seen in Table 2, overall, the ratio is less than unity in the 1996 census (0.93) and 1998 household members listing (0.88). The ratio by population group is also less than unity except among Indians in the 1998 household members listing. The ratios are questionable in the 1996 census data even among Indians and whites (0.88 and 0.87 respectively). The ratio in Sweden, Switzerland and United Kingdom (which are among the countries with the lowest fertility in the world) were respectively 1.0, 0.91 and 0.99 in 1995. A three-parameter stable population models chosen on the basis of fertility, mortality and rate of population growth in South Africa as estimated by Udjo (1998, 2003) for the period under consideration suggest a ratio of 1.11 for the population.

The stable population model suggests that underreporting of persons aged 0-4 relative to persons aged 5-9 at the time of the 1996 census may have been in the region of 18%. The bottom panel of Table 2 in comparison with the stable population model suggests that underreporting of the 0-4 age group relative to the 5-9 age group in Eastern Cape, the Free State and Limpopo provinces in the

20

1996 census was probably over 20%. The underreporting may be partly due to age shifting.

.

MODEL AGE DISTRIBUTIONS

The UN age-sex accuracy, Whipple's and Myers' (see, Shyrock, Siegel and Associates, 19760 blended indices are commonly used methods for evaluating the quality of repoted age-sex distributions. These are summary overall measures of quality, which do not provide detailed insights into the quality of the reports. Detailed insights can be gained from single or five-year population pyramids aided by demographic modelling. The demographic models constructed as further means of evaluating the age-sex distributions consisted of model age distributions for the different population groups on the basis of the smoothed 1970 census age-sex distributions which were then projected to 2001 on the basis of the fertility and mortality estimates provided by Udjo (1998, 2003) without the impact of migration and HIV/AIDS. Similarly, the smoothed 1970 census age-sex distributions by population group were projected incorporating migration and HIV/AIDS. Both sets of projections can be used to evaluate the quality of the reported 1996 and age-sex distributions from future surveys and censuses.

Appendices 1-2 show the differences between the logits of the observed and standard percentages under specified ages. The plots for Africans and coloureds

21

are more or less inverted C-shapted indicating the presence of errors in the reported male and female age distributions (female plots not shown) hence the need for smoothing. If there were no errors, the points would form almost a straight line. On the other hand, the plots for Indians and whites are almost a straight line except at the extreme old ages (Appendices 3-4) suggesting better quality of the age-sex reports (female plots not shown) among Indians and whites. The almost horizontal line in the plots for whites except at the extreme old ages (Appendix 4) suggests that the age distribution of whites was alreadyonsidered approximately stable in 1970.

The reported and smoothed age distributions shown in Appendices 5-8. Note that the smoothed age distributions do not fit very well at the two oldest age groups among whites (Appendix 8) but this is of very little significance as these age groups numerically, constitute less than 3% of the white population and are either near or at the terminal ages at which persons exit the population.

Appendices 9-13 compare the smoothed five-year age-sex distributions projected to 2001, with the reported age-sex distributions in the 1996 census by population group and a summary of the results in broad age groups is given in Table 3. As seen in the table 3, the 1996 model sex ratios are higher than corresponding sex ratios reported and adjusted by statistics South Africa (see bottom panel of Table 3, columns 3, 5 and 7). The difference arises largely from the sex ratio of the 15-

64 age group in the African population and to a lesser extent, the coloured population (see first two panels of Table 3, columns 3, 5 and 7).

In general, the reported and Statistics South Africa's adjusted 1996 census results suggest apparent deficit of males relative to females in the African and coloured populations that is inconsistent with the parameter estimates of demographic processes among Africans and coloureds as seen from the results of the model estimates.

Also, the reported and Statistics South Africa's adjusted 1996 census appear to understate the 0-4 age group as the proportion is inconsistent with the parameter estimates of demographic processes of the population. Among Africans for example, the ratio of the 0-4 and 5-9 age groups in 1996 was about 0.96 in the adjusted results of the Census, the present model suggests a ratio of 1.10 while a figure of 1.02 was derived from the OHS98 birth history. A simulation model also suggests that if present levels and trends in fertility and mortality continued marked indentation at the base of the pyramid of the African population would not occur before 2010. The model estimates for 2001 for the parameters described above are similar to those for 1996 and thus may may be used in evaluating the 2001 census results as discussed in the last session.

EFFECTS OF HIV/AIDS AND MIGRATION

Theoretically, the effects of HIV/AIDS and positive net migration on age distributions operate in opposite directions. While AIDS deaths in populations with generalised HIV epidemic reduce the relative proportions in certain age groups large scale positive net migration should increase the relative proportions in certain age groups. Furthermore, because HIV sero- prevalence is generally, higher for females than males in populations characterised by heterosexual transmission (such as in South Africa), this should lower age-specific sex ratios in certain age groups and also the overall sex ratio in populations with high HIV prevalence at some future date. On the other hand, since migration is sex selective, positive net migration could counter the effect of HIV/AIDS on age-specific sex and overall sex ratios. Table 4 shows the effect of HIV/AIDS and migration on South Africa's age distributions. As seen in the table, there is no evidence from the model estimates to suggest that the combined effect of net migration and HIV/AIDS is altering sex ratios substantially in South Africa's population at present.

DISCUSSION AND CONCLUSION

An apparent deficit of males was noted in the reported and adjusted 1996 census results among Africans and coloureds that is inconsistent with: (1) the parameter estimates of demographic processes in the populations; (2) the OHS98 birth history reports. It could be argued from the results of this study that the apparent deficit males may have resulted from over statement of the number of females in household member listings in certain age groups. Several factors may be at play for the over statement of females: (1) Age shifting may be part of the explanationn. For example at face value, the first panel of Table 3 would suggest that virtually all Africans in 1996 were in the age group 0-64 as virtually less than 0.1 percent were reported aged 65 and above. On the contrary the model estimates suggest the about 3% of the African population were probably in the age group 65 and over implying that in the household member listing the age of these persons were shifted downward to the 15-64 age group. Although Statistics South Africa corrected for this shift (see top panel of Table 3, column 4) in the final result of the census, the direction of the correction is debatable for the following reasons. Presumably, women are more likely to understate their age and this should be taken into consideration. However, it would appear this was not given sufficient consideration in Statistics South Africa's adjustment as it only raised the ratio of the 15-64 age group from 90 to 0.91 compared with 97 from the model, and birth history of the 1998 survey.

25

(2) Over statement of females could also arise from mis-statement of sex in the household member listing and this may be culturally rooted. For example, a colleague has suggested that linguistically the masculine is some time interchanged with the "feminine" in day-to-day discourse. This author recalls his (anecdotal) experience in Botswana: some Tswana would sometime refer to him as *"ra"* and shortly after as *"ma"*. While the former terminology is usually used with reference to the male, the latter is supposed to refer to a woman. It is probable that the feminine is sometime used interchangeably for the masculine among Africans in South Africa. This requires further investigation.

The results from this study suggest that the age-sex distribution from the household listing of the October household surveys and the 1996 may not be reliable especially among the African population. The OHS98 birth history appears to provide a better indicator of sex ratios. It is not clear why the mothers would report more accurately "sex" in the birth histories as compared to "sex" in the household member listing.

Reported differences in the absolute number of males and females in absolute terms based on the household member listing in the October household surveys (now replaced by the general household survey), and the 1996 census need re-examination since such differences have implications for unadjusted indicators derived from such data. It probable for example that the pattern of age shifting noted above combined with probable exaggeration of the number of females in

the 15-64 age group (i.e. in the labour market) may have to some extent, biased previous female and overall unemployment "rates" upward as the "total number of cases existing at a particular point in time" (i.e. unemployed) may have been exaggerated.

The results of this study also suggest that the 0-4 age group is understated in Statistics South Africa's data sets including the final results of the 1996 census. In the light of this, reliance on these data for certain kinds of planning (for example, in the education sector) is likely to be mis-leading.

In view of the above, caution should be exercised in producing estimates of male-female differences in indicators from the 2001 census and future household surveys. The above estimates provide a valuable tool for evaluating the age-sex distributions from the 2001 census results. As seen from the last two panels of Table 4, certain aspects of the age-sex distributions from the 2001 census are inconsistent with the model estimates. Notably, the ratio of the 0-4 and 5-9 age groups is lower in Statistics South Africa's adjusted 2001 census figures compared to the model estimates among Africans, Indians and whites. This suggests underestimation of the 0-4 age group in Statistics South Africa's adjusted figures. Also, the model estimates appear to suggest that overall sex ratios in Statistics South Africa's figures are too low among Africans and coloureds. These seemingly low overall sex ratios are attributable to the apparent low sex ratios of Africans and coloureds in the labour market.

In the past, the low sex ratios from Statistics South Africa's data sets have often been interpreted as indicating "missing males". This apparent "missing males" is probably due to overstatement of females in household member listing in the 1996 census. Since the 1996 census final results have been used for deriving weights for sampling and subsequently weighting of survey data, the error in sex ratios are carried over to the survey data (and indicators derived from them) since sex is either used as an explicit or implicit stratification variable in sampling and weighting. Lastly, the combined effect of net migration and HIV/AIDS appear to have very little impact on overall sex ratios presently in South Africa.

ACKNOWLEDGEMENT

REFERENCES

Bah S. (1999). A comparative assessment of independent sources of migration data: the case of the South African residents and citizens who have emigrated to five leading destination countries. Memeograph, Statistics South Africa, Pretoria.

Brunk M. (1996). Undocumented migration to South Africa: more questions than answers. IDASSA, Cape Town.

Crush J. (1999). The dimensions of irregularity in post-apartheid South Africa. International migration, Blackwell publishers, Oxford.

Dorrington R. (1999). To count or to model that is not the question: some possible deficiencies with the 1996 census results. Paper presented at the Arminel Roundtable workshop on the 1996 South Africa Census, Hogsback, 9-11 April.

Hirschowitz R. and Orkin M. (1996). Living in South Africa: Findings from the October 1995 household survey. Central Statistical Service, Pretoria.

Kaplan D., Meyer J., and Mercy B. (1999).). Brain drain: new data, new options. In, Trade and Industry Monitor, September, pp. 10-13.

McDonald D, Mashike L., and Golden C. (1999). The lives and times of African migrants and immigrants in post-apartheid South Africa. Migration policy series no. 13, IDASSA, Cape Town.

Phillips H. E., Anderson B.A., and Tsebe P. (1999). Sex ratios in South African census cata 1970 – 1996. Paper presented at the Workshop on Phase 2 of Census 1996 Review on Behalf of the Statistical Council, Johannesburg 3-4 December.

Sadie J. L. (1988). A reconstruction and projection of demographic movement in South Africa.

Sadie J. L. (1999). The missing millions. Paper presented to NEDLAC meeting. Johannesburg.

Shell R. (1999). An investigation into the reported sex composition of the South African population according to the census of 1996. Paper presented at the Workshop on Phase 2 of Census 1996 review, Wanderers Club, Johannesburg, 3-4 December.

Shryock H. S., Siegel J. S., and Associates (1976). The methods and materials of demography. Academic Press, New York, pp. 113-144.

Statistics South Africa (1999). The people of South Africa, population census 1996: primary tables, the country as a whole. Statistics South Africa, Pretoria.

Udjo E. O. (1997). Fertility and mortality trends in South Africa: the evidence from the 1995 October household survey, and implications on population projections. Central Statistical Service, Pretoria.

Udjo E. O. (1998). Additional evidence regarding fertility and mortality trends in South Africa and implications for population projections. Statistics South Africa, Pretoria.

Udjo E. O. (1999). Comment on R. Dorrington's "To count or to model that is not the question". Paper presented at the Arminel Roundtable Workshop on the 1996 South African Census, Hogsback, 9-11 April.

Udjo E. O. (2003). A re-examination of levels and differential in fertility in South Africa from recent evidence. Journal of Biosocial Science, vol. 35, pp. 413-431.

Table 1: Reported sex ratios in broad age groups in 1998 and 1996 by population group

0-64 age group	1970 census	OHS98 household members listing	OHS98 birth history	1996 census
African	0.975	0.883	0.993	0.930
Coloured	0.97	0.933	1.045	0.949
Indian	0.99	0.980	1.173	0.967
White	1.017	0.973	1.063	0.984
Total	0.983	0.898	1.009	0.938

0-14 age group	1970 census	OHS98 household members listing	OHS98 birth history	1996 census
African	0.991	0.998	1.014	0.987
Coloured	0.993	1.124	1.141	1.005
Indian	0.995	1.044	1.066	1.026
White	1.042	1.013	1.011	1.037
Total	0.998	1.011	1.028	0.993

15-64 age group	1970 census	OHS98 household members listing	OHS98 birth history	1996 census
African	0.963	0.816	0.966	0.897
Coloured	0.949	0.853	0.946	0.921
Indian	0.986	0.959	1.271	0.945
White	1.004	0.961	1.133	0.968
Total	0.972	0.838	0.984	0.909

Table 2: Ratio of 0-4 to 5-9 age group by population group and province

Population group

	1996 census	OHS98 household members listing	OHS98 birth History
African	0.933	0.874	1.015
Coloured	0.944	0.908	1.001
Indian	0.876	1.000	‹ 1.008
White	0.862	0.910	0.925
Total	**0.928**	**0.881**	**1.008**

(Ratio of 0-4 to 5-9 InStable population = 1.114)

Province

	1996 census	OHS98 household members listing	OHS98 birth History
Western Cape	0.977	0.908	0.941
Eastern Cape	0.865	0.858	1.023
Northern Cape	0.932	0.817	0.955
Free State	0.890	0.827	0.888
Kwazulu-Natal	0.938	0.963	1.099
North West	0.938	0.870	1.031
Gauteng	1.048	0.973	1.052
Mpumalanga	0.929	0.867	1.054
Limpopo	0.882	0.823	0.947

Table 3: Reported and model age distributions without HIV/AIDS and migration. 1996 and 2001

AFRICAN

Age group	1996 CENSUS Reported ratio/proportion	M/F	1996 CENSUS STATS SA ADJUSTED ratio/proportion	M/F	1996 CENSUS Model Ratio/proportion	M/F	2001 CENSUS Model ratio/proportion	M/F
0-4/5-9	0.934	-	0.958		1.101		1.107	
0-14	0.364	0.987	0.365	0.987	0.371	1.012	0.363	1.013
15-64	0.636	0.897	0.593	0.905	0.599	0.971	0.605	0.974
65+	0.000	0.600	0.042	0.601	0.030	0.663	0.032	0.660
Total	1.000	0.913	1.0000	0.919	1.000	0.975	1.000	0.976

COLOURED

Age group	1996 CENSUS Reported ratio/proportion	M/F	1996 CENSUS STATS SA ADJUSTED ratio/proportion	M/F	1996 CENSUS Model Ratio/proportion	M/F	2001 CENSUS Model ratio/proportion	M/F
0-4/5-9	0.944		0.964		1.058		0.976	
0-14	0.328	1.005	0.3330	1.006	0.331	1.024	0.316	1.025
15-64	0.634	0.921	0.6309	0.923	0.637	0.984	0.648	0.986
65+	0.037	0.678	0.0361	0.679	0.032	0.671	0.036	0.670
Total	1.000	0.937	1.0000	0.939	1.000	0.985	1.000	0.984

INDIAN

Age group	1996 CENSUS Reported ratio/proportion	M/F	1996 CENSUS STATS SA ADJUSTED ratio/proportion	M/F	1996 CENSUS Model Ratio/proportion	M/F	2001 CENSUS Model ratio/proportion	M/F
0-4/5-9	0.876		0.905		0.984		1.007	
0-14	0.274	1.026	0.2765	1.026	0.297	1.022	0.285	1.023
15-64	0.688	0.945	0.6854	0.946	0.658	0.990	0.664	0.989
65+	0.039	0.772	0.0381	0.772	0.045	0.821	0.051	0.807
Total	1.000	0.959	1.0000	0.960	1.000	0.991	1.000	0.988

WHITE

Age group	1996 CENSUS Reported ratio/proportion	M/F	1996 CENSUS STATS SA ADJUSTED ratio/proportion	M/F	1996 CENSUS Model Ratio/proportion	M/F	2001 CENSUS Model Ratio/proportion	M/F
0-4/5-9	0.866		0.874		0.990		0.968	
0-14	0.212	1.037	0.214	1.038	0.231	1.028	0.220	1.024
15-64	0.682	0.968	0.682	0.970	0.677	0.990	0.682	0.989
65+	0.106	0.685	0.104	0.686	0.092	0.745	0.098	0.734
Total	1.000	0.948	1.000	0.949	1.000	0.973	1.000	0.968

TOTAL

Age group	1996 CENSUS Reported ratio/proportion	M/F	1996 CENSUS STATS SA ADJUSTED ratio/proportion	M/F	1996 CENSUS Model Ratio/proportion	M/F	2001 CENSUS Model ratio/proportion	M/F
0-4/5-9	0.929		0.952		1.085		1.083	
0-14	0.341	0.993	0.343	0.993	0.348	1.015	0.340	1.015
15-64	0.609	0.909	0.609	0.916	0.614	0.976	0.619	0.977
65+	0.050	0.628	0.048	0.629	0.039	0.692	0.041	0.685
Total	1.000	0.920	1.000	0.925	1.000	0.976	1.000	0.976

Table 4: Statistics South Africa Adjusted and model age-sex distributions, taking account of migration and HIV/AIDS by population group in 1996 and 2001

Stats SA adjusted 1996

Age	African ratio/prop	M/F	Coloured ratio/prop	M/F	Indian ratio/prop	M/F	White ratio/prop	M/F
0-4/5-9	0.958		0.964		0.905		0.874	
0-14	0.365	0.987	0.333	1.005	0.277	1.026	0.214	1.038
15-64	0.593	0.905	0.631	0.923	0.685	0.946	0.682	0.970
65+	0.042	0.601	0.036	0.679	0.038	0.772	0.104	0.686
Total	1.000	0.919	1.000	0.939	1.000	0.960	1.000	0.949

Model, 1996

Age	African ratio/prop	M/F	Coloured ratio/prop	M/F	Indian ratio/prop	M/F	White ratio/prop	M/F
0-4/5-9	1.101		1.058		0.984		0.990	
0-14	0.371	1.012	0.331	1.024	0.297	1.022	0.231	1.028
15-64	0.599	0.971	0.637	0.984	0.658	0.990	0.677	0.990
65+	0.030	0.663	0.032	0.671	0.045	0.821	0.092	0.745
Total	1.000	0.975	1.000	0.985	1.000	0.991	1.000	0.973

Stats SA adjusted, 2001

Age	African ratio/prop	M/F	Coloured ratio/prop	M/F	Indian ratio/prop	M/F	White ratio/prop	M/F
0-4/5-9	0.918		0.957		0.843		0.854	
0-14	0.340	0.991	0.308	1.007	0.235	1.024	0.190	1.042
15-64	0.616	0.903	0.652	0.909	0.719	0.952	0.699	0.956
65+	0.043	0.530	0.040	0.647	0.046	0.712	0.111	0.708
Total	1.000	0.911	1.000	0.926	1.000	0.956	1.000	0.940

Model, 2001

Age	African ratio/prop	M/F	Coloured ratio/prop	M/F	Indian ratio/prop	M/F	White ratio/prop	M/F
0-4/5-9	1.008		0.954		1.002		0.930	
0-14	0.356	1.013	0.304	1.025	0.284	1.022	0.219	1.026
15-64	0.610	0.977	0.661	1.096	0.664	0.987	0.681	0.976
65+	0.034	0.669	0.035	0.670	0.051	0.808	0.100	0.726
Total	1.000	0.977	1.000	1.055	1.000	0.987	1.000	0.958

Figure 1: Reported sex ratios 0-64 age group in 1998 and 1996 by province

- OHS98 household members list
- OHS98 birth history
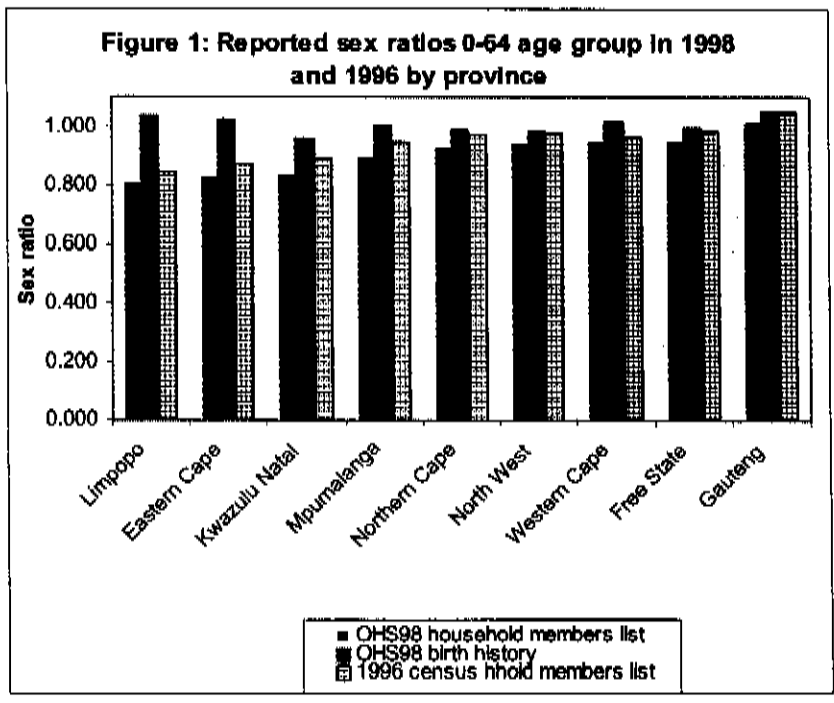- 1996 census hhold members list

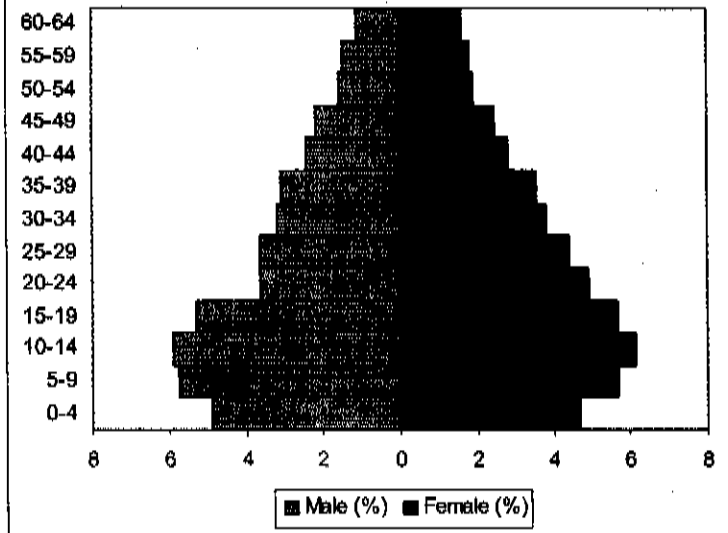**Figure 2: Age-sex distribution from the October Household Survey, 1995**

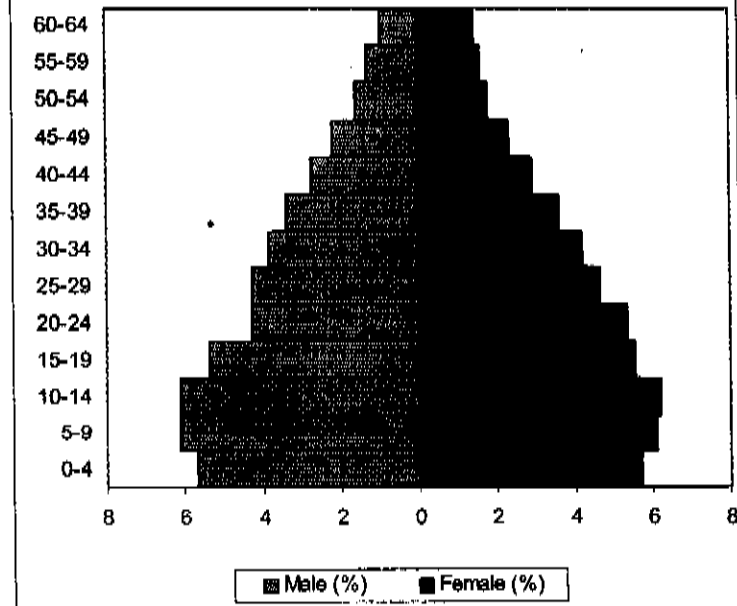**Figure 3: Age-sex distribution from the 1996 census**

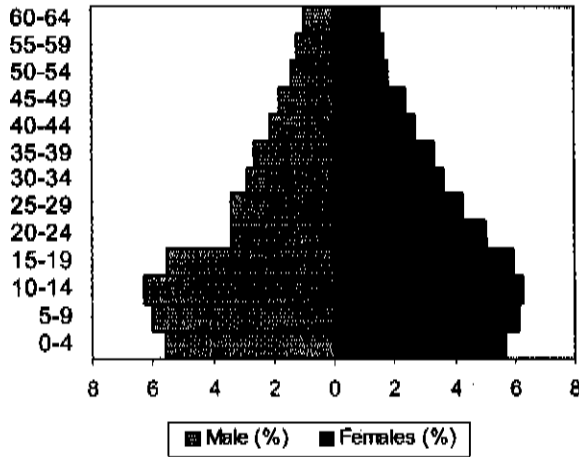Figure 4: Age-sex distribution from the October Household Survey 1997

Figure 5: Age-sex distribution from the Special Retrospective Survey, 1997
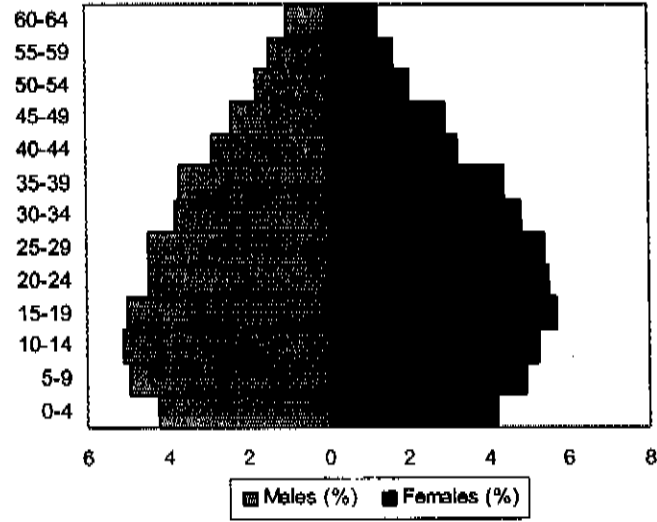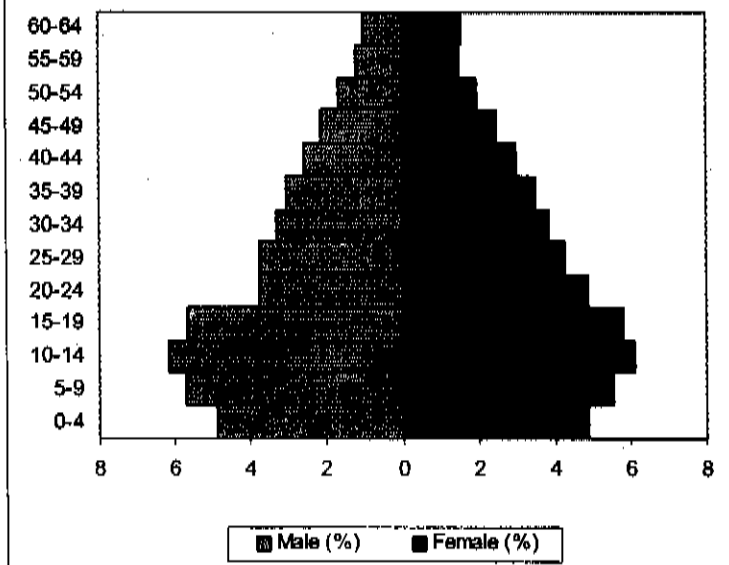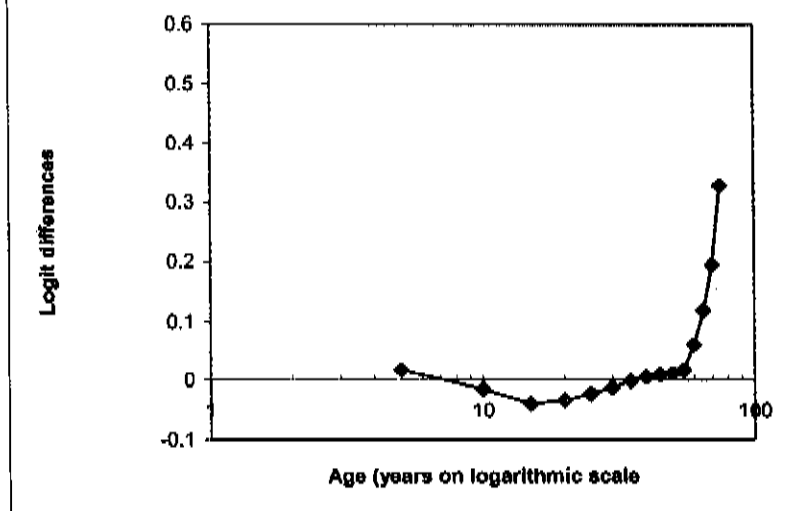
Figure 6: Age-sex distribution from the Income and Expenditure Survey, 2000

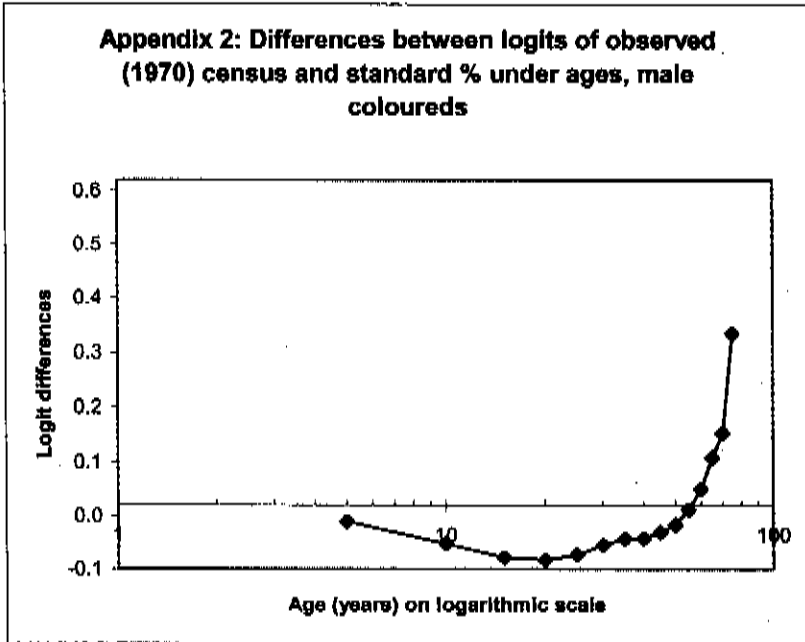## Appendix 1: Differences between logtis of observed (1970) census and standard % under ages, male Africans

Source: 1970 census

Appendix 2: Differences between logits of observed (1970) census and standard % under ages, male coloureds

Source: 1970 census

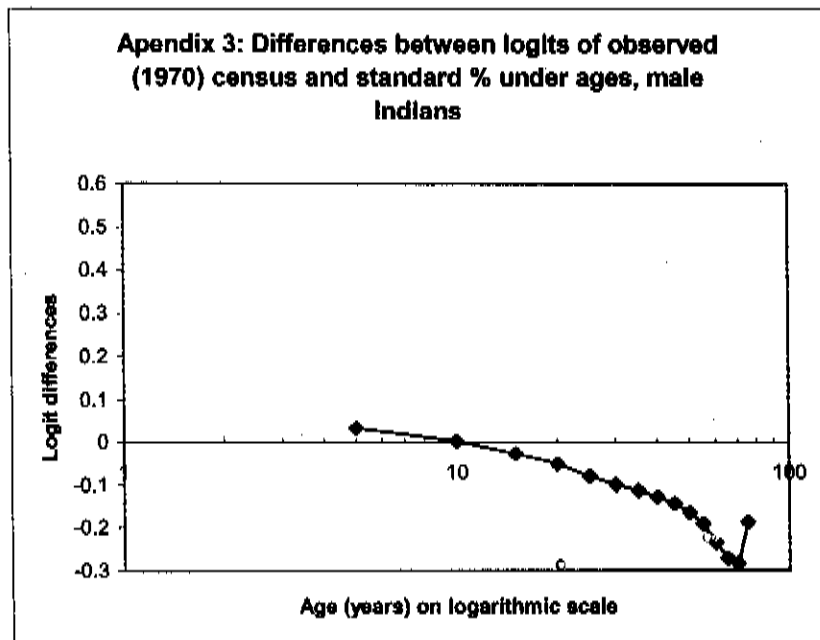**Apendix 3: Differences between logits of observed (1970) census and standard % under ages, male Indians**

Source: 1970 census

40

Appendix 4: Differences between logits of observed (1970) census and standard % under ages, male Whites

Source: 1970 census

Appendix 5: Reported and smoothed male age distribution, 1970 census: Africans
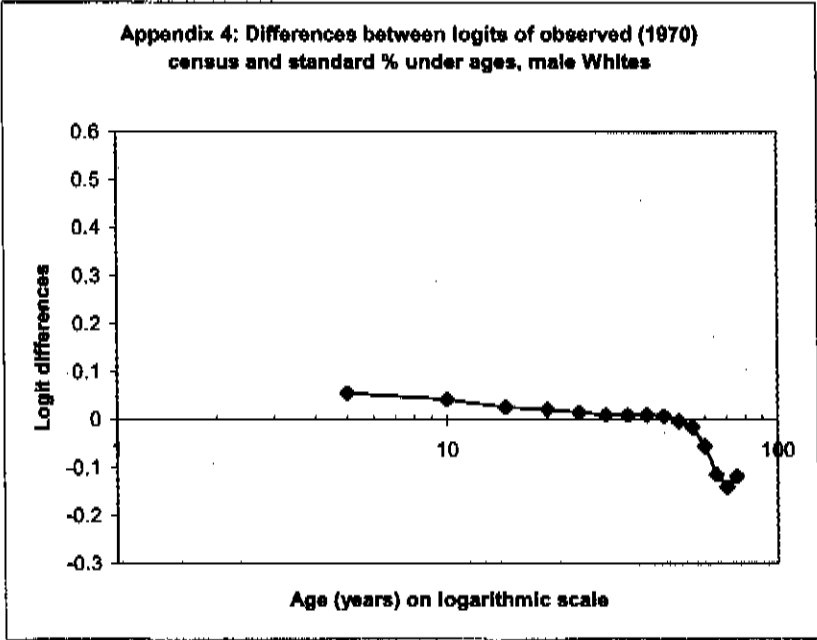
Source: 1970 census

Source: 1970 census

Appendix 6: Reported and smoothed male age distribution, 1970 census: Coloureds

Source: 1970 census

Appendix 7: Reported and smoothed male age distribution, 1970 census: Indians
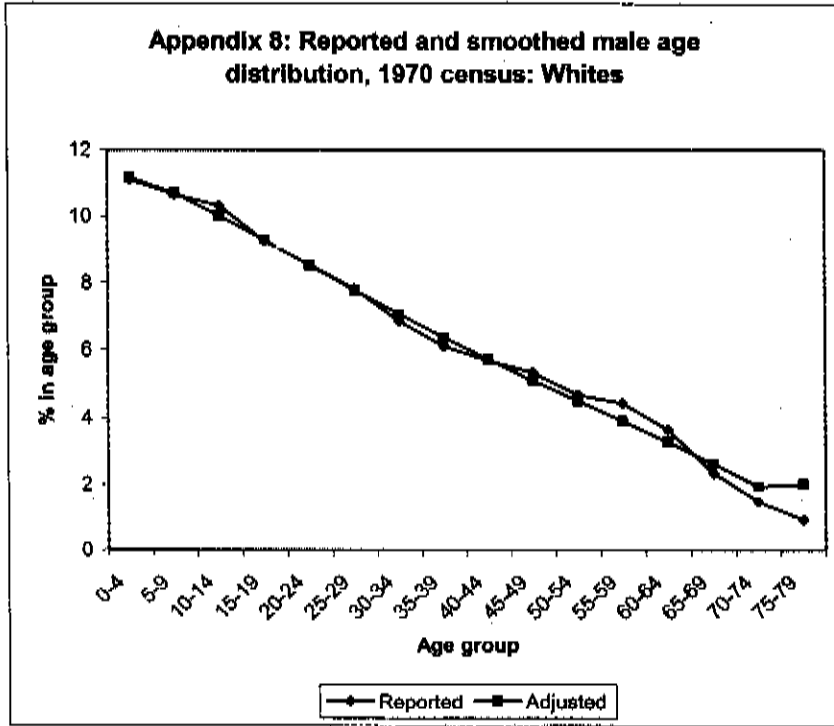
Appendix 8: Reported and smoothed male age distribution, 1970 census: Whites

Appendix 9: Reported and estimated sex ratios for
1970, 1996 and 2001 without HIV/AIDS and migration
AFRICAN

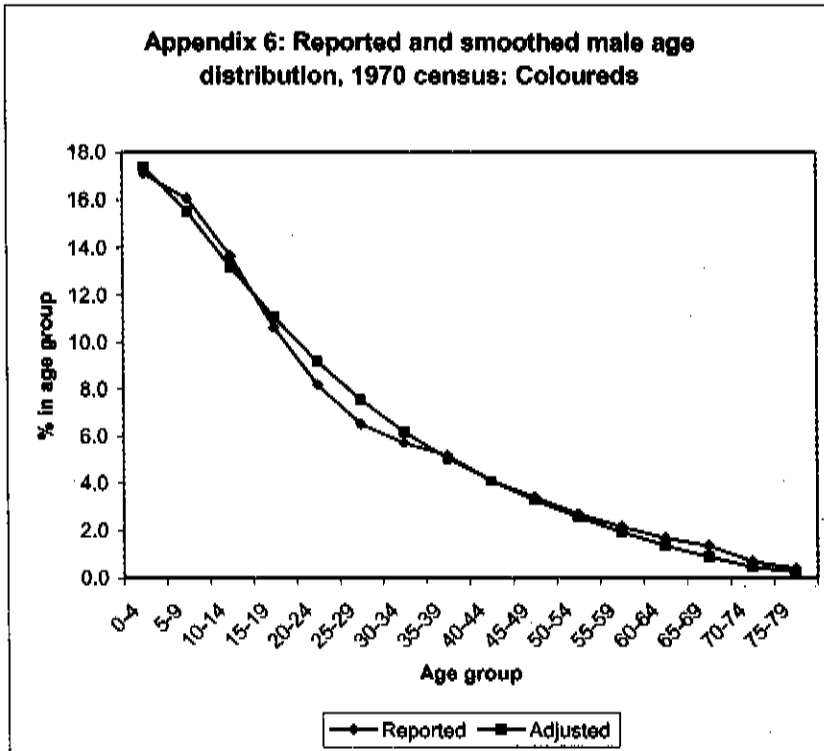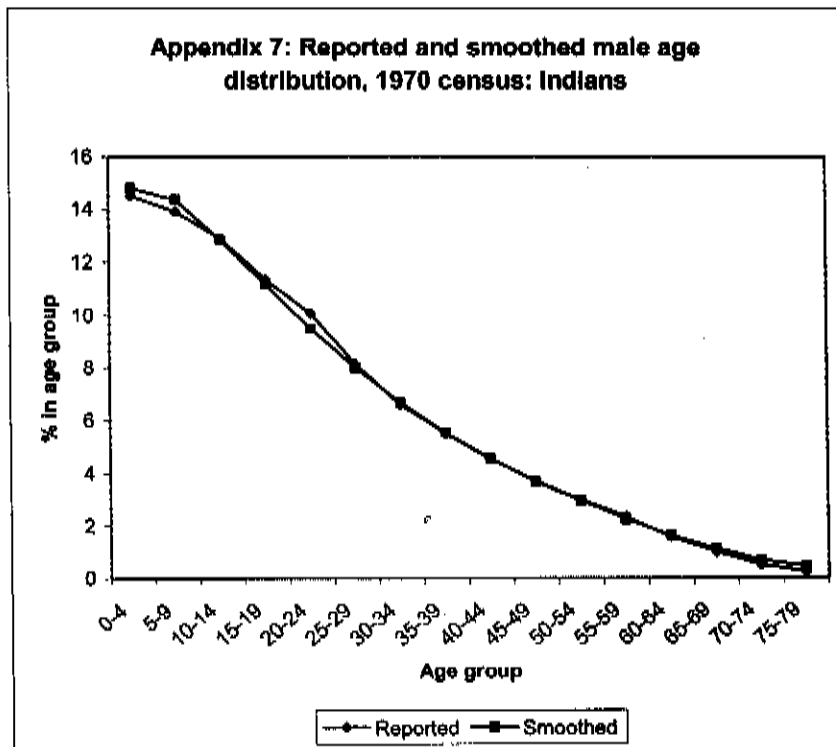| Age group | 1970 Census | | 1996 Census | | 2001 |
| | Reported | Estimated | Reported | Estimated | Estimated |
	M/F	M/F	M/F	M/F	M/F
0-4	98.8	99.2	99.2	101.5	101.5
5-9	99.7	101.0	99.4	101.3	101.3
10-14	98.9	100.7	97.6	100.9	101.1
15-19	94.1	99.8	94.7	100.4	100.6
20-24	96.4	98.8	90.4	99.8	100.1
25-29	96.6	97.6	90.4	97.8	99.5
30-34	94.3	96.5	89.4	99.2	97.5
35-39	97.7	95.3	93.3	99.0	98.7
40-44	99.2	94.2	92.4	97.5	98.0
45-49	106.2	93.0	93.5	95.0	95.8
50-54	97.8	91.8	87.5	91.4	92.4
55-59	101.9	90.6	77.9	86.5	87.6
60-64	81.0	89.4	57.5	80.2	81.1
65-69	83.8	88.0	61.0	72.8	73.2
70-74	73.5	86.5	65.3	65.3	65.1
75-79	74.1	84.0	58.6	59.3	58.5
80+	61.9	77.6	51.6	56.5	54.3
Total	**96.7**	**97.8**	**91.3**	**97.5**	**97.6**

Appendix 10: Reported and estimated sex ratios for
1970, 1996 and 2001 without HIV/AIDS and migration

COLOURED

Age group	1970 Census Reported M/F	1970 Census Estimated M/F	1996 Census Reported M/F	1996 Census Estimated M/F	2001 Estimated M/F
0-4	99.0	100.8	100.3	102.7	102.7
5-9	99.3	102.7	101.2	102.4	102.5
10-14	99.7	102.2	100.0	102.1	102.3
15-19	96.5	101.1	100.1	101.6	101.9
20-24	90.3	99.8	95.9	101.1	101.4
25-29	93.8	98.4	93.0	99.4	100.8
30-34	93.7	97.0	91.3	101.0	99.1
35-39	96.5	95.7	91.1	100.6	100.5
40-44	95.5	94.4	88.5	98.9	99.7
45-49	102.3	93.1	88.7	96.3	97.3
50-54	99.9	91.8	88.6	92.6	93.8
55-59	98.8	90.5	86.9	87.7	88.9
60-64	87.5	89.1	76.5	81.3	82.4
65-69	92.7	87.6	78.4	73.8	74.4
70-74	84.7	85.8	68.6	66.0	66.1
75-79	81.6	82.9	62.2	59.8	59.2
80+	65.1	76.0	46.9	56.5	54.5
Total	96.6	99.1	93.7	98.5	98.4

Appendix 11: Reported and estimated sex ratios for
1970, 1996 and 2001 without HIV/AIDS and migration

INDIAN

| Age group | 1970 Census | | 1996 Census | | 2001 |
| | Reported | Estimated | Reported | Estimated | Estimated |
	M/F	M/F	M/F	M/F	M/F
0-4	97.5	102.7	101.3	102.5	102.5
5-9	99.6	102.6	103.2	102.2	102.2
10-14	101.7	102.0	102.9	101.9	102.0
15-19	93.7	101.4	100.2	101.6	101.7
20-24	93.8	100.7	99.0	101.1	101.2
25-29	95.3	100.3	96.8	101.1	100.7
30-34	100.8	100.1	92.3	101.0	100.6
35-39	98.7	99.9	93.5	100.0	100.4
40-44	104.0	100.0	91.7	98.7	99.2
45-49	102.7	100.2	93.9	97.1	97.4
50-54	112.1	100.8	93.7	95.1	95.2
55-59	111.7	101.7	89.5	92.5	92.4
60-64	118.5	103.4	79.0	89.1	88.9
65-69	117.3	105.6	80.6	85.2	84.6
70-74	116.9	109.6	78.3	81.1	80.0
75-79	115.7	116.1	74.5	78.3	76.4
80+	118.2	89.8	64.8	79.3	75.8
Total	99.3	101.6	95.9	99.1	98.8

Appendix 12: Reported and estimated sex ratios for 1970, 1996 and 2001 without HIV/AIDS and migration

WHITE

Age group	1970 Census Reported M/F	1970 Census Estimated M/F	1996 Census Reported M/F	1996 Census Estimated M/F	2001 Estimated M/F
0-4	103.9	104.4	102.9	102.7	102.7
5-9	104.3	104.5	103.8	102.5	102.3
10-14	104.6	103.6	104.3	103.1	102.3
15-19	103.7	102.5	103.8	102.1	102.8
20-24	100.4	101.3	98.0	102.3	101.7
25-29	103.5	100.1	96.2	102.2	101.8
30-34	103.5	99.0	95.3	102.5	101.6
35-39	103.0	97.9	96.0	101.3	101.7
40-44	100.6	96.8	96.3	99.5	100.2
45-49	100.0	95.8	97.6	97.4	98.0
50-54	96.8	94.8	96.5	94.6	95.1
55-59	96.5	93.8	95.1	91.0	91.4
60-64	89.9	92.9	89.6	86.4	86.5
65-69	85.0	92.0	83.3	81.0	80.7
70-74	71.0	91.1	74.1	75.0	74.3
75-79	63.5	89.5	63.6	69.8	68.5
80+	56.8	84.7	46.5	67.0	64.3
Total	**99.5**	**99.6**	**94.8**	**97.3**	**96.8**

Appendix 13: Reported and estimated sex ratios for
1970, 1996 and 2001 without HIV/AIDS and migration

TOTAL

| | 1996 Census | | 2001 |
| | Reported | Estimated | Estimated |
	M/F	M/F	M/F
0-4	99.6	101.7	101.7
5-9	99.9	101.5	101.5
10-14	98.4	101.2	101.3
15-19	96.0	100.7	100.9
20-24	91.7	100.2	100.4
25-29	91.4	98.6	99.9
30-34	90.3	99.9	98.3
35-39	93.4	99.5	99.3
40-44	92.5	98.0	98.5
45-49	93.8	95.6	96.3
50-54	89.6	92.2	93.1
55-59	82.3	87.7	88.6
60-64	65.2	81.9	82.6
65-69	66.8	75.3	75.5
70-74	68.0	68.5	68.1
75-79	60.2	63.2	62.1
80+	50.0	61.0	58.4
Total	**92.0**	**97.6**	**97.6**