ADULT MORTALITY IN SOUTHERN AFRICA USING DEATHS REPORTED BY HOUSEHOLDS: SOME METHODOLOGICAL ISSUES AND RESULTS

Rob Dorrington[a], Ian Timæus[b] and Simon Gregson[c]

a   Centre for Actuarial Research, University of Cape Town
b   Centre for Population Studies, London School of Hygiene & Tropical Medicine
c   Imperial College, London

Censuses in developing countries quite often ask households to report on the number of deaths in a period immediately prior to the census but these data don't seem to have been used much to produce estimates of adult mortality in Southern Africa in recent years. This paper analyses the potential biases in these data and applies a combination of the generalized growth balance and synthetic extinct generations methods to data adapted from censuses in Zimbabwe to produces estimates of mortality. These estimates are compared with those produced by other researchers and found to be broadly consistent and the results are interpreted in the context of similar applications to data from Swaziland and Botswana.

1.   Introduction

Despite being collected in a number of censuses, deaths reported by households are often not used to estimate mortality in Southern Africa and where they are, they are assumed to be accurately reported and mortality rates derived directly. The reasons for this are not clear but probably stem from an uncertainty by those working for central statistical offices about how to estimate the completeness of reporting, together with the fact that early attempts to derive mortality estimates from these data were disappointing (Timæus 1991). In such situations the data are sometimes not even published.

This is a pity since South Africa and Zimbabwe are the only two countries in Southern Africa where it has been possible to make use of vital registration data to measure mortality ((Dorrington, Bradshaw and Wegner 1999; Dorrington, Moultrie and Timæus 2004; Feeney 2001; Timæus, Dorrington, Bradshaw *et al.* 2001)), and even then, in the case of Zimbabwe, only 30-40% of female deaths were registered, leading to significant uncertainty around the estimates of the mortality of women. Thus, with the 2000 round of censuses, deaths reported by households recorded in the census could be a vital source of information about mortality in recent years.

Manuals and textbooks tend to treat deaths reported by households as a source of data similar to vital registration data but there are reasons to suspect that these data are subject to different biases from those affecting vital registration data. In addition, these data usually only cover the year prior to the census so, unless one can assume either that the population is stable or that mortality rates are constant, one needs to adapt the existing methods for evaluating the completeness of death reporting in an intercensal interval to derive mortality rates from these data.

2.   Potential biases in data on deaths reported by households

Standard texts (such as UN Population Division(1983; 2002)) on estimating incompleteness of reporting of deaths all mention deaths reported by households as a potential source of data on deaths, however, it is treated as synonymous with vital registration data, with little or no attention being paid to how biases in these data might differ. Although a number of authors (Centre for Population Studies 1994; Hill, Choi and Timæus 2005 in press; Timæus 1993) have mentioned potential sources of bias in deaths reported by households there doesn't appear to have been any investigation into the likely extent of these biases.

Two of the underlying assumptions of the methods for estimating completeness of reporting of deaths are that they represent the population being investigated and that completeness of reporting is the same for all ages (or at least the age range of interest). Clearly deaths reported by households can only represent the mortality of the population living in households. However, in most countries with poor vital registration systems by far the majority of the population live in households and thus these estimates may be considered to be representative of the population as a whole if a representative sample or a census is taken of all the households in a country.

Thus the major concern is whether the potential biases in such data might lead to a violation of the assumption that reporting is equally complete by age. In this regard the potential problems with these data are:

- Under-reporting (as a result of: failure to report a recent death; confusion around the length of the reference period; non-coverage of specific areas/populations; disintegration of the household on the death; or, for completeness, the institutionalising of segments of the population (e.g. the aged))
- Over –reporting (as a result of: confusion around the length of the reference period; or people seen as belonging to more than one household)

- Age misreporting (either age heaping or age exaggeration).

Some of these errors, such as the failure to report a recent death, confusion over the reference period and age heaping are unlikely to violate the assumption significantly and will be corrected for through the estimate of completeness of the reported deaths. Others, in particular households disintegrating on the death and age exaggeration at the older ages, have the potential to violate the assumption that completeness of reporting is constant with respect to age, and will need to be investigated, both for effect and for significance.

Vital registration data, in general, can be expected to be more representative and less biased than data from households with the exception possibly of bias caused by differential coverage of specific populations (either population groups or regions, such as urban and rural), which could prove to be significant. In some more developed settings registration of male deaths may be more complete than that of female deaths due to the existence of life assurance cover associated with employment.

Investigating the extent of these biases is of course difficult, since in most such settings where one would be trying to use deaths reported by households to estimate mortality rates one does not have benchmark mortality rates against which to assess the accuracy of the rates being produced. One could make use of DSS sites to test the quality of reporting of these data in these specific populations by comparing the answers to the questions about deaths in the last year to the records of the household, but we are not aware of any such studies. On the other hand, where there is an alternative means of deriving national estimates, as is the case in South Africa, using the vital registration data, one can compare the number of deaths reported by household against the number expected on the basis of the alternative estimates of mortality. Such a comparison was made by Dorrington, Moultrie and Timæus (2004) using data from the 2001 census and a sample of death registrations. A reworking of these figures using the full sample registration data is shown in Figure 1.

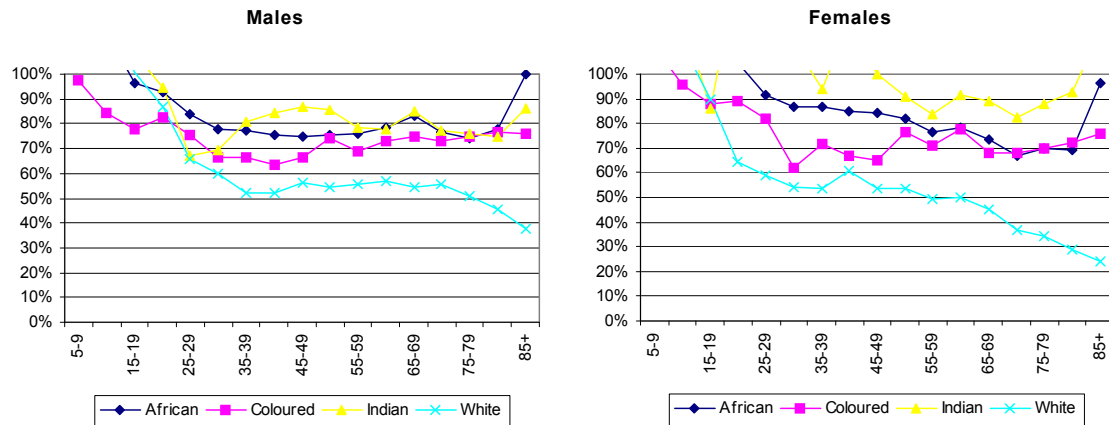|	Males	Females

Figure 1: Estimates of completeness of reporting of deaths by households by population group in the 2001 census

A number of interesting features are apparent from Figure 1. The first is, bearing in mind that by far the majority of the population are African, that with the exception of the White population, completeness of the reporting of male adult deaths is largely constant with respect to age whereas the completeness of reporting of female deaths falls with age for most adult ages. This difference between men and women probably reflects the high proportion of households headed by women that disintegrate on the death of that woman, which in part will be due to wives surviving their husbands.

The second is the clear exaggeration of age at death of African men and women as evidenced by the big increase in completeness for those in the open age interval (85+). This is not evident in the other population groups.

The third is the different levels of completeness for the different population groups. In particular the fact that completeness of reporting of deaths by the Coloured (mixed races) population is notably less than that of Africans, and that of White households significantly below that of the other population groups. Although the reasons for the lower completeness of the reporting by Coloured households is not immediately clear, that of White households is probably due to a generally low participation of the White population in the census, emigration of households, households disintegrating on death of one of two members and much higher institutional residence at older ages.

In summary though, these results suggest that the biases, certainly as they manifest in South Africa, are unlikely to lead to a significant violation of the assumption of constant completeness of reporting by age in the case of men, but may lead to some violation in the case of women. However, any violation that produces a known pattern of completeness

estimates by age can be allowed for, to some extent, in the adjustment of the mortality rates for incompleteness of reporting.

3. Methods for estimating the completeness of reporting of deaths by households

When it was originally proposed that one might use deaths reported by households to estimate mortality, it was envisaged that the Growth Balance method, devised by Brass (1975), be used to estimate the completeness of the reporting. This method assumes, *inter alia*, that the population is stable, which at the time, was a reasonable assumption for many countries to which the method might have been applied. However, today, this is no longer a reasonable assumption to make and where possible one of the variable-*r* methods (such as the generalised growth balance (GGB) method (Hill 1987) or the synthetic extinct generations (SEG) method (Bennett and Horiuchi 1981; 1984)) should be applied. The application of both of these methods is described in detail in the United Nations manual on estimating adult mortality (United Nations Population Division 2002).

Essentially, the GGB method relies on the balancing equation applied to the population of persons over a given age. Thus, for a population closed to migration:

$$N(x) - \left[P_2(x+) - P_1(x+) \right] = D(x+)$$

where $P_1(x+)$ and $P_2(x+)$ represent the numbers of persons aged x and over in the population at the first and second censuses respectively, $D(x+)$ represents the number of deaths during the intercensal period to persons aged x and over, and $N(x)$ represents the number of persons reaching exact age x during the intercensal period.

Dividing through by the number of person years lived during the intercensal period by persons aged x and over produces the following relationship:

$$n(x) - r(x+) = d(x+)$$

which when the true quantities are replaced with the observed quantities[*] allows us to estimate not only the extent of completeness of the reporting of deaths, but also the completeness of one census relative to the other. It is a simple matter to adapt this relationship to take net

---

[*] In other words $P_i(x+) = P_i^o(x+) / k_i$ and $D(x+) = D^o(x+) / c$ where P^o and D^o are observed numbers and k_i and c are measures of completeness, assumed to be constant over all (adult) ages.

migration into account, simply by adding the annual rate of net immigration to the right hand side of the above equation.

Although a description of the SEG method entails moderately complicated formulae, the basic idea, originally proposed by Carrier (1958), is simple, namely that where there is no error caused by lack of completeness in the data, the number in the population at a specific age at a point in time must be equal to the number of deaths in future years arising out of that cohort. These numbers of deaths in turn can be estimated, on the assumption that mortality remains constant over time, from the number of deaths recorded in an interval by noting that deaths at any particular age will grow at the population growth rate at that age. A comparison of the this estimate with the estimate of the population at that age derived from the censuses, thus gives an estimate of the completeness of reporting of the deaths.

In practice, apart from data on the number of people and the number of deaths reported by households in one census, one is likely to have data on the number of people from the preceding census and possibly even data on the number of deaths reported by households from that census as well. In order to apply one of the variable-r methods one needs estimates of the number of people by age at two time points and the number of deaths between those points, which is a problem when censuses are five or ten years apart and deaths are reported for only a year or so.

The better of these two situations is where one has two censuses in which each asked households to report on the number of deaths. In this case it is suggested that one estimate the number of deaths in the intercensal period by assuming that the number of deaths at each age grew exponentially over the relevant period.

Where only the later census asked households to report on the number of deaths then the best that one can do is to use the census populations to estimate the number of people in the population a year before the second census and the reported deaths to estimate the number of deaths occurring in the year before the second census. Since one is effectively using the average annual intercensal growth rate as an estimate of the growth rate in the last year of this period, one could effectively produce the same estimate of mortality by estimating the number of deaths in the intercensal period by multiplying the estimated number deaths in the 12 months prior to the second census by the length of the intercensal period and proceed as above. However, it is important to realise that in this case the rates thus produced are applicable in the year prior to the second census and are not an average for the whole period.

However, where all the data one has is the single census where households report on the number of deaths over a specified period, it is recommended that the Growth Balance method be applied as a diagnostic tool if for no other reason, rather than simply either assuming that the data are 100% complete, or that they are useless. The plot of the partial birth rates against the partial death rates, and the estimated population growth rate, give some indication of the quality of the data, which in turn might give some idea of whether the data are in need of some correction for completeness or not.

The methods as originally presented, and most often used, are based on the assumption that the population is closed to migration. However, where net migration may have been significant one should adapt these methods to allow for this. The problem, though, is that in these countries one rarely has reliable estimates of the net migration rates by sex and age, and one might have to allow for migrants by deciding on an age distribution (either by using a standard such as that from Rogers and Castro (1981) or by using information on the change in the stock of the number of immigrants from the country in question captured in the main receiving country or countries) and some estimate of overall net number of migrants (either from a reconciliation of the census populations or by aggregating the numbers received by the main receiving countries where these are thought to be accurate). An alternative would be to set the overall number of migrants so as to minimize the deviation from the fitted line in the GGB method, however, if the pattern of migration by age is similar to that of the deaths, such as may well be the case where there are significant AIDS deaths, this might not produce a very good estimate of under-reporting of deaths, as one might have trouble distinguishing out-migration from under-reported deaths.

4.   Application to Zimbabwe

Recent mortality rates for Zimbabwe needed to be estimated to help decide whether an apparent decline in HIV prevalence in that country was due to increased mortality due to AIDS (UNAIDS 2005). Mortality rates were estimated by applying a combination of the GGB and SEG methods suggested by Hill and others (2005 in press). In this procedure the GGB is used to correct for any undercount of one census relative to the other, and then SEG method is applied to the corrected estimates of population numbers and the number of deaths reported by the households to estimate the completeness of reporting of deaths by households and hence estimate adult mortality rates.

We had available population numbers from three censuses (1982, 1992 and 2002) and deaths reported by households from the 1992 and 2002 censuses.

Since there has been substantial emigration of Zimbabweans to South Africa over the past five to ten years it was thought necessary to adapt the method to take into account an allowance for net emigration over the two periods, the 1992 to 2002 intercensal period and the year prior to the 2002 census. It was assumed that prior to the 1992 census net migration was negligible, most post-civil war migration being over by 1982. Migration post 1992 was derived iteratively by first deciding on a rough order of magnitude of total migration for the period and age distribution by differencing the projected population from 1992 based on the estimate of mortality assuming no migration, and assuming no migration for ages above 55 for men and 50 for women. This was then used to revise the estimate of mortality and a new age distribution of migration derived, and so on until the proportions of migrants in each age group stabilised. The rough order of magnitude of total migration was then altered to see if the fit to the GGB regression could be improved by significant changes to this number.

This application to the 1992-2002 period produced a surprisingly good fit for men and an acceptably good fit for women, with estimates of completeness of reporting of 92% for men and 84% for women, as shown in the figures below. However, in the case of women there is a distinctive falling off of completeness with age which suggests that the data could be subject to bias due to the disintegration of households on death of (older) women. This could mean that a better estimate of the mortality of women could be produced by assuming a declining level of completeness as age increases, starting maybe above 90% at the younger ages and falling to 70% at advanced ages, rather than using an average of 84% for all adult ages.
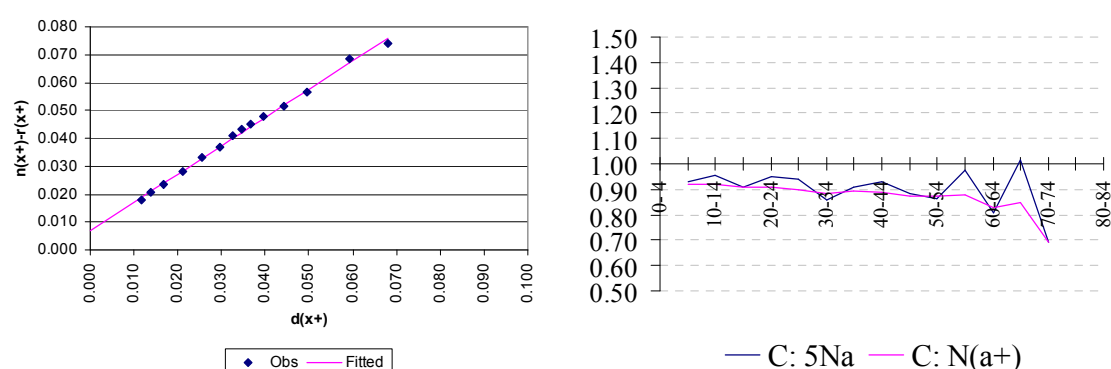

Figure 2: GGB and SEG plots for Zimbabwean male deaths reported by the 1992 and 2002 censuses
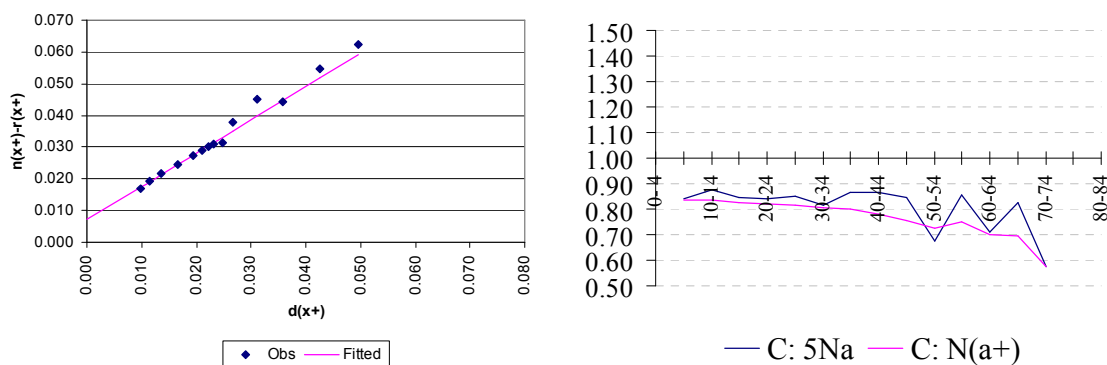
Figure 3: GGB and SEG plots for Zimbabwean female deaths reported by the 1992 and 2002 censuses

The estimation of completeness for the year preceding each of the censuses, separately, suggest that the data for the year prior to the 1992 census were 110% and 106% complete for men and women respectively, and for the year preceding the 2002 census, 95% and 70% complete, respectively. Naturally these estimates, being derived from a single census reporting on deaths by households, are less robust but the comparison below suggests that the estimates for 1992 are not out of line with estimates derived by other demographers using different data sources. The estimate for 2002 is a little less convincing, particularly for women, and after inspection it might be decided that there is no reason (given the results of previous years) for there to be such a big difference between the completeness of men and women, and perhaps an average of the two estimates of completeness might be a more reasonable estimate of the level of completeness.

Figure 4 compares the rates derived from household deaths with those derived from vital registration data by Feeney (2001) and from the DHS data on mortality of siblings by Timæus and Jasseh (2004). Also included on the graph are estimates of adult mortality from Manicaland for the period 1998-2003 (plotted at the mid-point) (Lopman, Barnabas, Hallett *et al.* 2006). Both the estimates for 1992 and the 1992-2002 period (plotted as the mid-point between the two censuses) are very consistent with the other estimates, while the estimates for 2002 and for the 1982-1992 period based on vital registration data, particularly for women, look out of line with the trends of the rates from the other sources.
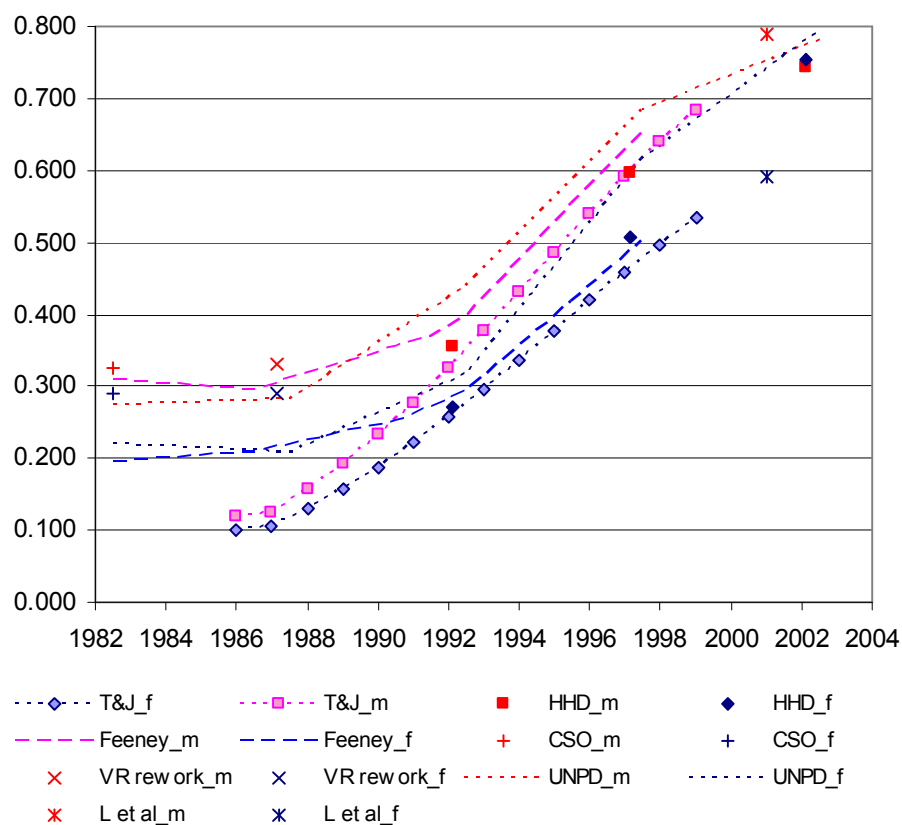
Figure 4: Comparison of adult mortality (45q15) from various sources
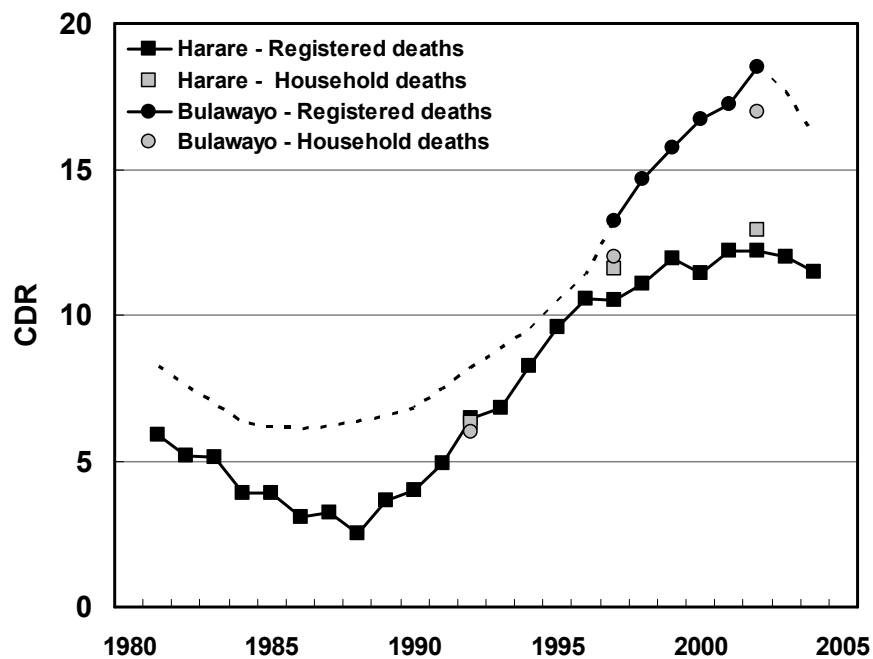

Figure 5: Comparison of implied crude death rates (CDRs) for Bulawayo and Harare, implied by applying estimates of completeness to CDRs derived from deaths reported by households

As a further test of the estimates of completeness produced by this research they were applied to crude mortality rates (CDRs) derived directly from deaths reported by households in Harare and Bulawayo and the these results compared to estimates derived using vital registration data for these cities (Figure 5). The results of this comparison suggest that the combined completeness of reporting of male and female deaths is broadly consistent in all cases except Bulawayo, 1992. On the other hand, comparison of the rates with those for Manicaland suggest that the estimate of completeness for females may be too low, however, for various reasons, these estimates need to be treated with some caution (due to the size of the sample, extended period over which base and follow up surveys are carried out, the exclusion of households lost to follow up and the fact that these are not national estimates).

Interestingly the GGB plus SEG combination method was applied to the vital registration data used by Feeney for the period 1982 – 1992. These estimates are included on Figure 4 and show that while the estimates for the men are broadly consistent with those derived by Feeney (using a variant of the GGB method only) this is not the case for the women, where the rate we have derived is significantly higher (i.e. estimated completeness significantly lower) than that produced by Feeney. Also included on Figure 4 are estimates of adult mortality produced by the Zimbabwe CSO (1985) for 1982 using a completely different method (child mortality to determine alpha, the level of Brass's General Standard). These estimates are quite consistent with those produced using the combination of GGB and SEG and might suggest, along with the gap between male and female siblinghood rates, that female mortality may have been somewhat closer to male mortality in the 1980s than estimates produced by Feeney, and other evidence, suggest. As might be expected the siblinghood data are prone to underestimate mortality going back in time.

5.   Applications to other countries

Estimation of completeness of reporting of deaths reported by households in censuses for Swaziland and Botswana have met with similar success as can be seen from the comparisons shown in Figures 6 and 7.

In the case of Swaziland (Figure 6) the estimates of adult mortality produced using deaths reported by households (HHD) are compared to those published by the Swaziland CSO (SCSO) (Swaziland Central Statistical Office 1991, 2002), implied by the UNDP projections (United Nations Population Division 2005), and by one of the authors as part of earlier

exercise using orhphanhood data to decide on a suitable level of the South Princeton standard table (Coale, Demeny and Vaughan 1983) (CARe).
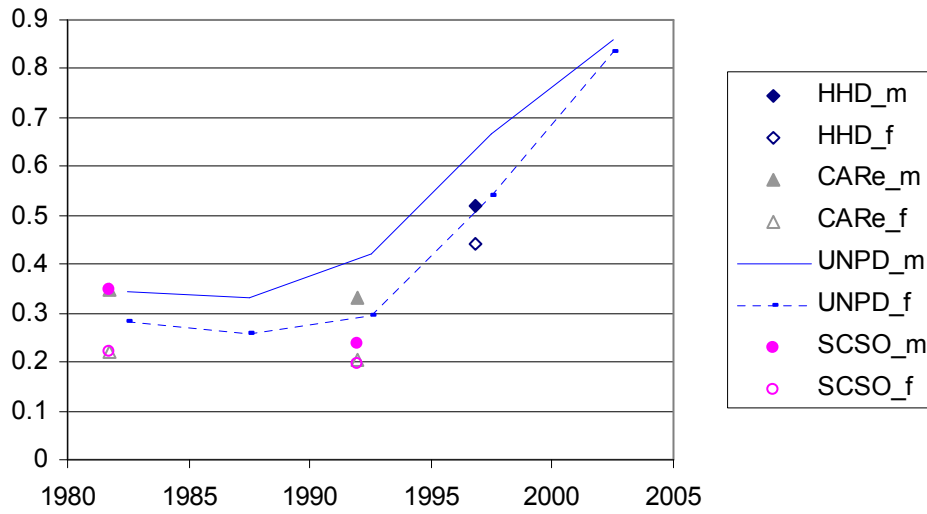

Figure 6: Comparison of adult mortality (45q15) from various sources – Swaziland

As far as the estimates of adult mortality in Botswana are concerned, these have been compared to estimates published by the Botswana CSO (BCSO91 for the 1991 estimates (produced by Eric Udjo) and BCSO01 for the 2001 estimates (produced by Rolang Majelantle) (Botswana Central Statistical Office 1987, 2004), Timæus (1993), and those implied by the UNDP projections (United Nations 2005).
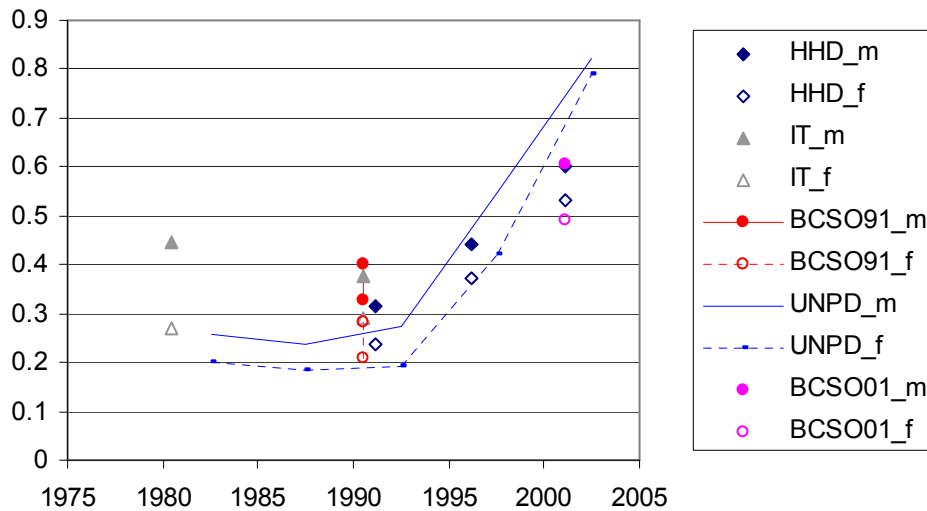

Figure 7: Comparison of adult mortality (45q15) from various sources – Botswana

From these two alternative applications we see that where we are able to compare the results against estimates derived from other data sources the estimates derived using deaths reported by households do not look at all unreasonable. (In this regard it should be pointed out that the BCSO estimate for 2001 also made use of the deaths reported by households, deriving the rates directly, implicitly assuming that reporting was 100% complete.)

In addition, all three case studies suggest that mortality rates underlying the most recent UN Population Division population projections rise too steeply, probably as a result of the assumption of too high a prevalence in these countries.

6. Discussion

Comparison of the deaths reported by households with the expected number on the basis of vital registration data corrected for under-reporting suggest that completeness of deaths reported by households may not be independent of age, since it is a function of the respondent's concept of household, whether the household disappears on the death and what proportion (particularly of the elderly) live in institutions. However, in many countries these distortions may well not be significant and where they exist, may well be detected and hence allowed for to some extent through a diagnosis of the plots produced as part of the methods to estimate completeness.

Unlike the experience in the 1970s and early 1980s investigation of the estimates produced using deaths reported by households in the three southern African countries considered as part of this research suggest that where one has fairly reliable estimates of the census population (and where significant, estimates of the level and possibly the pattern of migration) one is able to produce reasonable estimates of adult mortality. However, to some extent this conclusion may be coloured by the impact of HIV/AIDS on mortality which is so significant that it tends to swamp biases in the adjusted empirical estimates, which would make it difficult, otherwise, to track the trend in mortality rates.

This research has concentrated on censuses as the source of data. However these questions are often asked as part of household demographic surveys, as opposed to censuses. The usefulness of these data as a means of estimating mortality still needs to be investigated thoroughly, but preliminary investigations suggest they may be more problematic than censuses possibly due the sample not adequately representing the population against which completeness is being assessed. There is also the possibility that quality of information, such as the age of the deceased, is poorer in surveys than censuses.

Given the high levels of HIV prevalence being measured in many African countries, Southern African countries in particular, coupled with the uncertainty around these estimates, particularly where there are substantial differences between measures produced from antenatal clinic surveys and household surveys, establishing accurate estimates of the level of mortality is an important third measure of the extent of the epidemic. This research suggests that it is possible to produce useful estimates of mortality using data on deaths reported by households in the census, particularly if these data are available for two consecutive censuses. It also identifies the nature of biases that might be found in such data and suggests ways in which the data may be used with standard indirect techniques to produce estimates of mortality.

References

Bennett, N. G. and Horiuchi, S. 1981. "Estimating the completeness of death registration in a closed population", *Population Studies* **47**(2):207-221.

Bennett, N. G. and Horiuchi, S. 1984. "Mortality estimation from registered deaths in less developed countries", *Demography* **21**(2):217-233.

Botswana Central Statistical Office. 1987. *1981 Population and Housing Census, Analytical Report.* Gabarone: Central Statistics Office, Republic of Botswana. Available:

Botswana Central Statistical Office. 2004. *Analytical Report, 2001 Population and Housing Census.* Gabarone: Central Statistics Office, Republic of Botswana. Available:

Brass, W. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defected Data.* North Carolina: Carolina Population Center.

Carrier, N. H. 1958. "A note on the estimation of mortality and other population characteristics, given death by age", *Population Studies* **12**:149-163.

Centre for Population Studies, (Sloggett, A., W. Brass, S. M. Eldridge, I. M. Timæus, P. Ward, and B. Zaba). 1994. *Estimation of Demographic Parameters from Census Data.* Tokyo, Japan: United Nations Statistical Institute for Asia and the Pacific. Available:

Coale, A. J., Demeny, P. and Vaughan, B. 1983. *Regional Model Life Tables and Stable Populations.* New York: Academic Press.

Dorrington, R. E., Bradshaw, D. and Wegner, T. 1999. *Estimates of the Level and Shape of Mortaltiy Rates in South Africa Around 1985 and 1990 Derived by Applying Indirect Demographic Techniques to Reported Deaths.* Cape Town: Medical Research Counicl.

Dorrington, R. E., Moultrie, T. A. and Timæus, I. M. 2004. *Estimation of mortality using the South African 2001 census data.* Monograph 11. Centre for Actuarial Research, University of Cape Town. Available: http://www.commerce.uct.ac.za/care/Monographs/Monographs/Mono11.pdf

Feeney, G. 2001. "The Impact of HIV/AIDS on Adult Mortality in Zimbabwe", *Population and Development Review* **27**(4):771-780.

Hill, K., Choi, Y. and Timæus, I. M. 2005 in press. "Unconventional approaches to mortality estimation ", *Demographic Research* **S4**

Hill, K. H. 1987. "Estimating census and death registration completeness", *Asian and Pacific Population Forum* **1**(3):8-13.

Lopman, B. A., Barnabas, R., Hallett, T. D., Nyamukapa, C. *et al.* 2006. "Assessing adult mortality in HIV-1-afflicted Zimbabwe (1998-2003)", *Bulletin of the World Health Organization* **84**:189-197.

Rogers, A. and Castro, L. J. 1981. *Model Migration Schedules.* Laxenbeurg, Austria: International Istitute for Applied Systems Analysis.

Swaziland Central Statistical Office. 1991. *1986 Swaziland Population and Housing Census, Vol. 4 Analytical Report.* Mbabane: Central Statistics Office, Swaziland. Available:

Swaziland Central Statistical Office. 2002. *1997 Swaziland Population and Housing Census, Vol. 4 Analytical Report.* Mbabane: Central Statistics Office, Swaziland. Available:

Timæus, I. M. 1991. "Adult Mortality: Levels, Trends and Data Sources," in Feachem, R. G. and Jamison, D. T. (eds). *Disease and Mortality in Sub-Saharan Africa.* World Bank, pp. 87-100.

Timæus, I. M. 1993. "Adult mortality," in Foote, K. A., Hill, K. H. and Martin, L. G. (eds). *Demographic Change in Sub-Saharan Africa.* Washington, D.C.: National Academy Press, pp. 218-255.

Timæus, I. M., Dorrington, R., Bradshaw, D., Nannan, N. *et al.* 2001. "Adult mortality in South Africa, 1980-2000: From apartheid to AIDS," Paper presented at Population Association of America 2001 Annual Meeting. Washington DC, March 29-31 2001.

Timæus, I. M. and Jasseh, M. 2004. "Adult Mortality in Sub-Saharan Africa: Evidence from Demographic and Health Surveys", *Demography* **41**(4):757-772.

UNAIDS. 2005. *Evidence for HIV decline in Zimbabwe : a comprehensive review of the epidemiological data*. UNAIDS/05.26E. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS). Available:

United Nations. 1983. *Manual X: Indirect Techniques for Demographic Estimation*. New York: United Nations.

United Nations. 2005. *World Population Prospects: The 2004 Revision, Highlights*. New York: Department of Economic and Social Affairs.

United Nations Population Division. 2002. *Methods for Estimating Adult Mortality*. ESA/P/WP.175. New York: Department of Economic and Social Affairs. Available: www.un.org/population/publications/adultmort/complete.pdf

United Nations Population Division. 2005. *World Population Prospects: The 2004 Revision CD-Edition - Comprehensive Dataset*. New York: Department of Economic and Social Affairs, United Nations. Available:

Zimbabwe Central Statistics Office. 1985. *Main Demographic Features of the Population of Zimbabwe: An advance report based on a ten percent sample*. Harare: Central Statistics Office, Zimbabwe. Available: