CONTENTS

Demographic analysis of the population of Karonga District, northern Malawi, 1980–90 1
Tobias Chirwa, Basia Zaba, Sian Floyd, Simon Malema, Jorg Ponninghaus, Paul Fine

The estimation of infant mortality from proportions dying among births in the past 24 months 25
John Blacker and William Brass

Population mobility and household dynamics in rural South Africa: implications for demographic and health research 43
Victoria Hosegood, Justus Benzler, Geoff C Solarsh

ISSN 1682-4482
9771682448008

Southern African Journal of Demography

Vol 10(1&2) December 2005

Southern African Journal of Demography

Volume 10(1&2) December 2005

Southern African Journal of Demography

Volume 10(1&2)
December 2005

CONTENTS

Demographic analysis of the population of
Karonga District, northern Malawi, 1980–90. 1
Tobias Chirwa, Basia Zaba, Sian Floyd, Simon Malema, Jorg Ponninghaus and Paul Fine

The estimation of infant mortality from proportions dying among births
in the past 24 months 25
John Blacker and William Brass

Population mobility and household dynamics in rural South Africa:
implications for demographic and health research 43
Victoria Hosegood, Justus Benzler and Geoff C Solarsh

The *Southern African Journal of Demography*
ISSN: 1682-4482

EDITORS Professor Rob Dorrington and Dr Tom Moultrie
Centre for Actuarial Research, University of Cape Town

ADDRESS FOR CORRESPONDENCE
The Southern African Journal of Demography
Centre for Actuarial Research
University of Cape Town
Private Bag
7701 Rondebosch
SOUTH AFRICA

Fax: +27 21 689 7580
Email: sajd@commerce.uct.ac.za

DTP: User Friendly
Printed and bound by Creda Communications, Eppindust

Demographic analysis of the population of Karonga District, northern Malawi, 1980–90

Tobias Chirwa[1,2], Basia Zaba[3], Sian Floyd[2], Simon Malema[4], Jorg Ponninghaus[4], Paul Fine[2]

Abstract

Total population surveys were carried out in 1980–84 (LEP-1) and 1986–89 (LEP-2) in Karonga District, northern Malawi, as part of a major cohort study of leprosy and tuberculosis. Retrospective information on fertility was collected from adult women in selected villages. Mortality estimates for the adult population were based on reported inter-survey deaths. This paper describes the demographic dynamics of the population in this time period, which coincided with the start of the HIV epidemic.

A total of 112 886 individuals were interviewed in LEP-1, of whom 88 544 were also examined in LEP-2. Of this population, 46 per cent were under 15 years of age and about 20 per cent could not give a precise year of birth. Procedures for smoothing the age distribution and correcting the under-ascertainment of infants are explained.

Total fertility in the interval between the two surveys was 6.4 children per woman and the crude birth rate was 48 per thousand. The crude death rate for the same interval was about 18 per thousand, and the growth rate was 4 per cent per year. Infant mortality was around 120 per thousand live births, under-5 mortality was around 190, the index of adult mortality (proportion of those who survived to age 15 dying before age 60) was 37 per cent, and life expectancy was around 50 years. Analyses adjusted for age, sex and socio-economic factors show higher mortality in north compared to south Karonga (rate ratio of 1.29, 95 per cent CI: 1.19, 1.38); and in those with estimated rather than precise years of birth (rate ratio of 1.14, 95 per cent CI: 1.03, 1.25).

The similarities of the estimates of the age–sex distributions, and summary measures of mortality and fertility to the 1987 Malawi census and 1992 Malawi Demographic and

---

1 Chancellor College, University of Malawi
2 Infectious Disease Epidemiology Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine
3 Centre for Population Studies, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT. (Corresponding author *basia.zaba@lshtm.ac.uk*)
4 Karonga Prevention Study, Chilumba, Malawi

Health Survey results mutually support the reliability of Karonga epidemiological survey data, and add details not available through the standard demographic survey and census methods, particularly with respect to adult mortality.

Keywords
Mortality, fertility, sex ratios

INTRODUCTION

Populations in southern Africa are experiencing a complicated demographic transition, with the world's highest prevalence of HIV infection occurring in areas with the world's highest crude fertility rates, and severe pressures on natural resources. The long-term demographic implications of this conjunction, though not yet clear, have potentially important effects on the social and economic structure of these populations (Lesthaeghe and Jolly 1992).

An unusual opportunity to examine the demography of a rural southern African population has arisen in the context of a large epidemiological project begun in 1979 in Karonga District, northern Malawi. Designed originally as a study of leprosy and tuberculosis, this project [known first as the LEPRA Evaluation Project (LEP) and later as the Karonga Prevention Study (KPS 2004)] carried out two total population surveys during the 1980s, following up more than 100 000 people (Ponnighaus, Fine, Bliss *et al.* 1987, 1993). From subsequent HIV testing of sera collected in these two surveys, and from later epidemiological surveillance, it is now known that HIV was introduced into this population before 1982, and prevalence increased to approximately 4 per cent by 1990 and 13 per cent by 2000 (Crampin, Glynn, Ngwira *et al.* 2003, Glynn, Ponninghaus, Crampin *et al.* 2001). The data covered by the two initial surveys thus describe the demographic profile of this population at the time of the early spread of HIV. A third total population survey, maintaining the individual linkages established in the earlier work, is currently underway, and will in due course provide a detailed description of the demographic impact of HIV in the population.

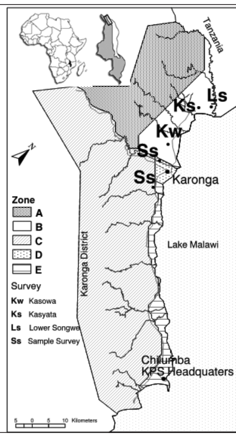
Karonga is a rural district whose economy is based almost entirely on subsistence agriculture and fishing in Lake Malawi. The district covers 3 355 square kilometres and had population densities of 36, 44 and 58 persons per square kilometre during the 1977, 1987 and 1998 census respectively. The annual growth rate as observed in 1977–87 and 1987–98 intercensal periods were 3.3 per cent and 2.5 per cent respectively. The languages mostly used for communication are Tumbuka and Nkhonde. The district consists predominantly of Christians and over 75 per cent of the population are literate. The sex ratio of about 93 shows more females resident within the district than males. The district had no easy land access route until 1980, when a road was built down the plateau escarpment south of the district. This led to a rapid increase in movement within the district and between Karonga and the rest of the country during the 1980s.

Several longitudinal demographic studies have been established in southern Africa in recent years (INDEPTH 2002), but the Karonga project began earlier and covered a larger population than any other, so that the recently established demographic surveillance system is unique in its historical antecedents. This paper describes the demography of the district population over the decade 1980–90. These early data provide baseline measurements of demographic and social patterns against which to examine what is currently happening as a consequence of the HIV epidemic. Above all, they provide estimates of adult mortality in the absence of HIV – information that is not readily available from other sources such as national censuses and household surveys.

METHODS

Demographic data were collected within a large epidemiological study of leprosy and tuberculosis in Karonga District, northern Malawi (Ponninghaus, Fine, Bliss *et al.* 1987). Two household surveys were conducted (LEP-1: 1980 to 1984 and LEP-2: 1986 to 1989), covering almost the entire district, excluding the southern tip in LEP-1 and the sparsely inhabited western hills in LEP-2. Additional surveys were carried out in groups of selected villages in 1984 and 1993–96 in which women were asked about the number of children they had ever borne, and how many were still alive.

Figure 1 Map showing Karonga District, divisions of ecological zones and special survey areas



The district was divided into five geographical "zones" on the basis of general ecological features (see Figure 1). These were the northern hills (Zone A), the northern lake shore area (Zone B), the southern hills (Zone C), the semi-urban area around the district capital (Zone D) and the southern lake shore area (Zone E). These zones were later grouped as Northern (A and B) and Southern (C, D and E) corresponding roughly to the main linguistic division in the area: Tumbuka in the south, Nkhonde in the north.

Details of the fieldwork methods have been described elsewhere (Ponninghaus, Fine, Bliss *et al.* 1987). For each individual, interviewers collected information on birth year, sex, mother and father's names and identification numbers, level of education, main occupation, village and household in which the individual was resident and position in household. Each individual was identified by an identification number (six digits plus an algebraically determined check digit) and initially assigned to the household in which they were first found and interviewed. Each household was assigned a unique (5-digit) "household number".

An effort was made in LEP-2 to trace the whereabouts and status of all individuals seen in LEP-1, to observe whether they died, changed household or migrated to other areas within or outside Karonga. Using this information, direct estimates of overall and age–sex specific mortality were obtained by calculating the person years at risk allowing for censoring due to loss to follow-up.

Among the risk factors for mortality that could be identified with the available data were housing quality, main occupation, level of education, geographical location of household, age and sex of individuals, and knowledge of birth year. These were investigated using a Poisson regression model. Housing quality was categorised in terms of construction material of dwellings i.e. houses made of burnt bricks, sun-dried bricks/pounded mud and grass temporary shelters.

Identification codes for the parents of each individual were recorded, so it was possible to link a woman to all individuals who had mentioned her as their mother, and to obtain direct estimates of surviving children resident in the area, even though standard demographic questions relating to child survival (David, Bisharat, Hill *et al.* 1990; Hill 1991) were only asked in selected villages in a separate survey. There were 55 790 women who were mentioned as a mother at least once during the survey period, approximately 60 per cent of whom were still living in the area at the time of the LEP-1 and LEP-2 surveys.

Special surveys were conducted in Kasowa and Lower Songwe areas (Figure 1) between 1993 and 1996, in which women were questioned regarding the number of live born children they had had, how many were still alive, and how many were resident in the area. Indirect estimates of child mortality were then computed for these areas, using the children ever born/ children surviving (CEB/CS) technique developed by Brass (David, Bisharat, Hill *et al.* 1990). Applying the age-specific CEB/CS ratios obtained in these areas to the surviving children of living mothers reported in LEP1 and LEP2 also enabled us to estimate mean CEB at the time of the earlier surveys.

For individuals who did not know their precise year of birth, this was estimated with reference to well-known local events (Ponninghaus, Fine, Bliss *et al.* 1987). The birth year assigned to these individuals was the mid-point between the date of the earliest event they recalled and the previous local calendar event. Since the event calendar featured just five events (occurring in 1914, 1934, 1946, 1958 and 1964), the assignment of birth dates to the mid-point of time intervals resulted in pronounced heaping of births in the years 1924, 1940, 1952 and 1961. This is reflected in age heaping that is slightly diffuse, as interviews were spread over several calendar years. Age was calculated by subtracting the assigned or precise year of birth from the year an individual was interviewed. If birthdays and interview dates were spread evenly throughout the year this would produce a half-year upward shift in estimated age – this was allowed for by assigning half the number in each single year of age (apart from infancy) to the preceding age.

To avoid age heaping as a result of birth date estimates based on the local events calendar, a procedure for smoothing the age distribution was employed (explained in the appendix). The smoothed age distribution was obtained for those with estimated birth dates by assigning them a random year of birth between the appropriate event calendar years, based on the birth year distribution of those with known years of birth within the respective intervals.

Field forms were checked, coded and entered onto computers in Malawi. Data are currently held in Oracle and Foxpro databases from which files are extracted for analysis. Analyses presented here were carried out in STATA version 7 (2001).

RESULTS

The total number of individuals seen in LEP-1 and LEP-2 surveys were 112 886 and 146 129 respectively. The population structure remained very similar in the two surveys, with 46 per cent under age 15, and 52 per cent female. If we assume that the average time between household contacts in the two surveys was six years, this would suggest a growth rate of 4.3 per cent, assuming similar levels of coverage and participation in the two rounds.

The largest upward adjustment for the half-year shift in calculating age last birthday occurred in infants, since there is no younger age group to compensate. Only 2 740 infants were originally recorded as age zero, but after transferring half of those originally reported as aged one, the total estimated number of infants was 5 157. The second largest upward adjustment was at age 55, a net gain of 441, and the largest downward adjustment occurred at age 59, where the net loss was 378.

Figure 2a shows the observed age distribution in LEP-1. The apparent age heaping for adults is a direct result of birth estimates based on the local events calendar. The smoothed age distribution is shown in Figure 2b.

Figure 3a shows the observed age distributions of males and females by five-year age groups. Clearly, the female age distribution is more distorted than that of males, particularly at ages over 25. This reporting pattern is largely attributable to the use of birth year estimates for individuals who could not report their actual year of birth. The proportion of individuals who did not know their year of birth increased with age, and was consistently higher for females than for males, with an overall figure of 30 per cent (17 479/58 264) for females and 11 per cent (5 716/51 964) for males. Twenty-one per cent and 15 per cent of the population did not report their year of birth in LEP-1 and LEP-2 respectively. Figure 3b shows the population age and sex distributions after adjustment for non-reporting of birth year.

Sex ratios

The sex ratios (male/female) were 0.904 and 0.936 during LEP-1 and LEP-2 respectively. Further evidence of the effect of the use of the event calendar on estimated age is shown in the graph of sex ratios, Figure 4a. Even though wider age groups are used after 50 to reduce this effect, the sex ratios for LEP-1 and LEP-2 reveal the same patterns except that the ratios at LEP-2 are shifted forward by approximately five years, the time between LEP-1 and LEP-2. After the re-assignment of birth years (Figure 4b), there is still some distortion between ages 25 and 50, and the increase in the sex ratio at older ages is still apparent. The anomalies are less pronounced in LEP-2.

Fertility

The structure of the epidemiological database formed from these surveys allows mother and child records to be linked – this record linkage goes much further than the usual "own children" method, since children who have

Figure 2a Age distribution of the population of Karonga District, northern Malawi identified in the 1979–84 survey (LEP-1)

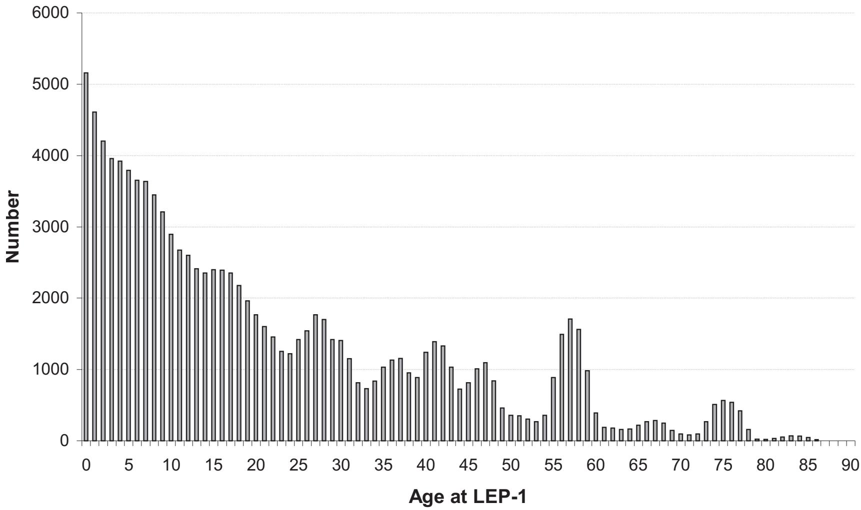


Figure 2b Age distribution after redistributing those with birth estimates in proportion to those with precise years of birth in LEP-1 (1979–84)

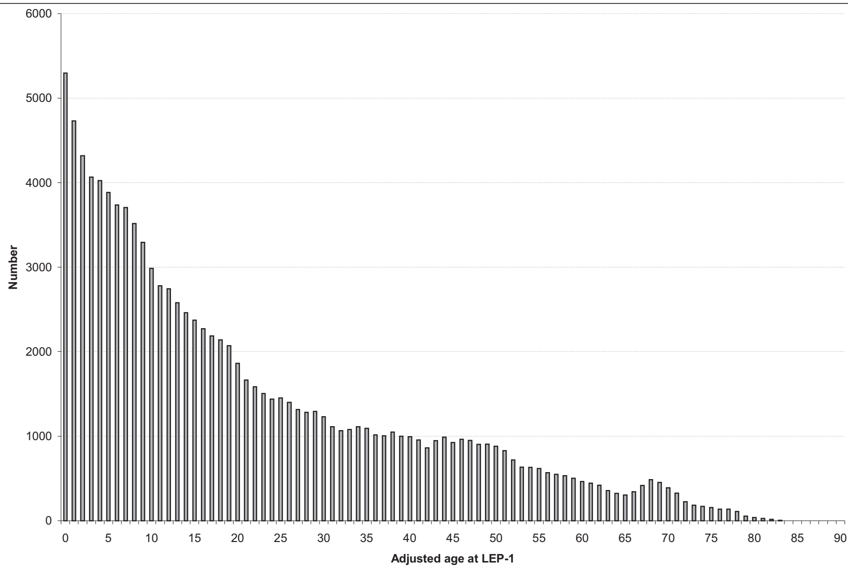


Figure 3a Population by age and sex for Karonga District, northern Malawi: LEP-1, 1979–84

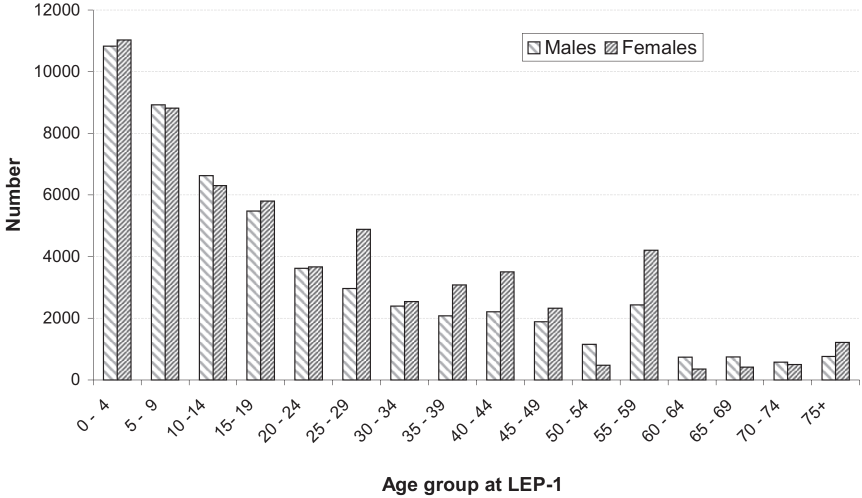


Figure 3b Population by age and sex, after redistributing those with birth estimates in proportion to those with precise years of birth in LEP-1 (1979–84)

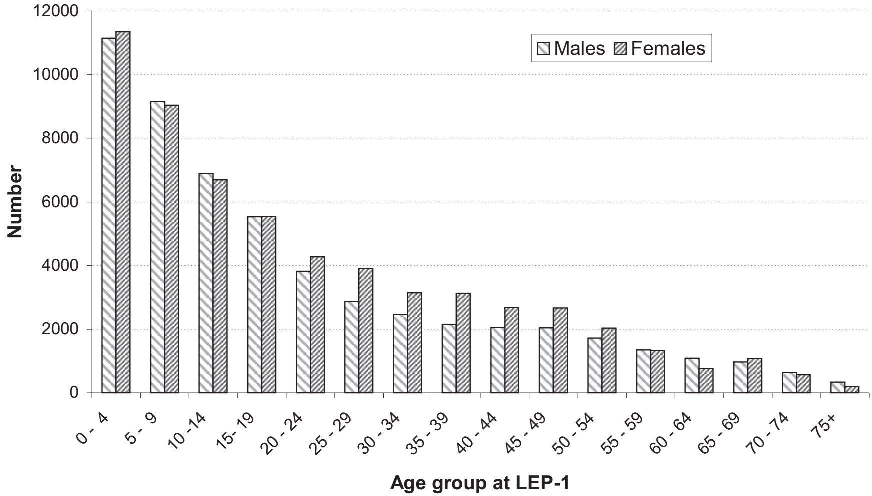


Figure 4a Age-specific sex ratios during LEP-1 and LEP-2 surveys. Karonga District, Northern Malawi, 1979–89

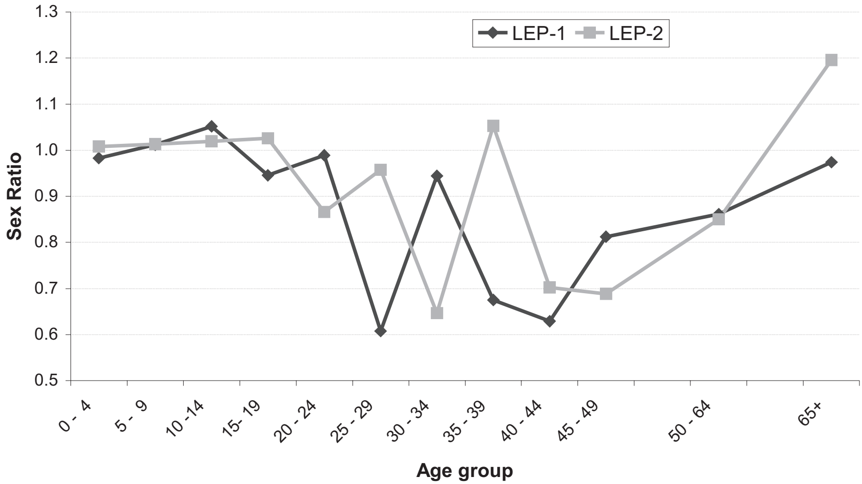
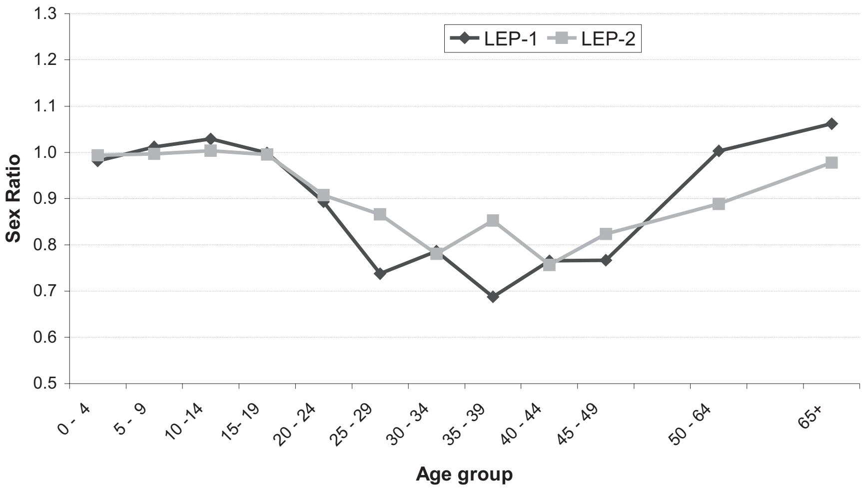


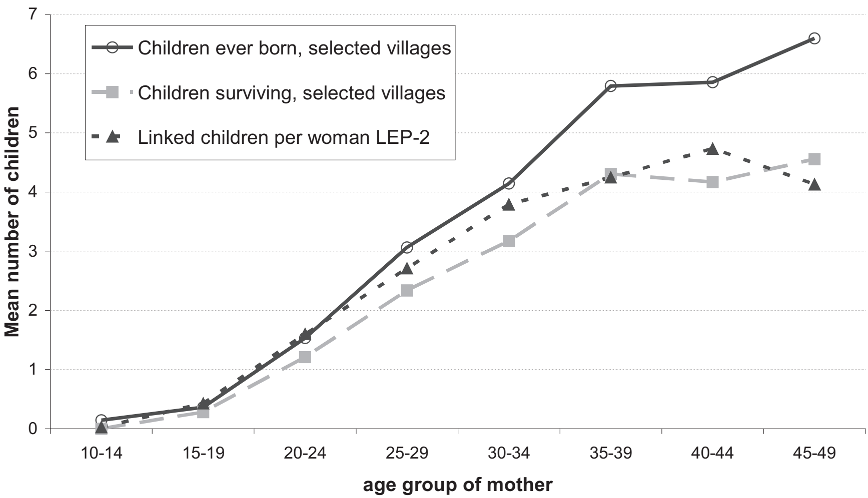
Figure 4b Age-specific sex ratios after redistributing those with birth estimates, LEP-1 and LEP-2 surveys



reached adulthood are included. It is possible to link children and mothers who live in different households, even if mothers or children left the district or died in the inter-survey interval, by using proxy reports of other close relatives. The records of 88 136 surviving children were linked to those of 25 468 mothers aged 15–49 and living in the district at the time of the LEP-2 survey. The number of surviving children identified per mother at LEP-2 ranged from 1 to 14, with an average of 3.5 surviving children per mother. The total number of women aged 15–49 at LEP-2 was 33 258, implying that 7 790 (23 per cent) of these women were childless, giving a mean number of surviving children per woman of 2.7. The mean surviving children per woman based on record linkage rose systematically with age, as expected, and this is shown in Figure 5a. The proportion childless by age 50 was just over 5 per cent.

This figure also shows the age pattern of mean children ever borne (CEB) and mean children surviving (CS) reported by women in the retrospective survey in selected villages in the North. The mean number of surviving children per woman in the age range 15–49 in the retrospective survey was 2.4. The trend in mean CS increases with age of mother and is similar to mean

Figure 5a Mean Children Ever Born (CEB) and mean Children Surviving (CS) in 1993–96 sample surveys, and mean linked children per woman at LEP-2 by age of mother, Karonga District, Malawi

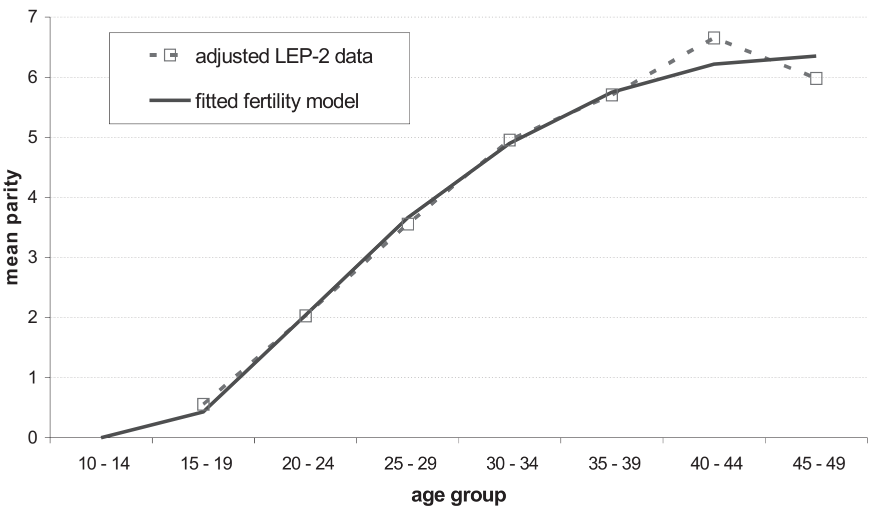


surviving children based on record linkage, justifying the use of the sample survey results to make proportionate adjustments to the record linkage data to allow for child mortality and thereby obtain parity estimates at the time of the LEP-2 survey. For example, mean CEB for women aged 45–49 in the sample survey was 6.6 and mean number of surviving children for women in this age group was 4.6. Since the mean number of linked children per woman at LEP-2 in this age group was 4.1, our best estimate of the CEB for this age group at LEP-2 is 5.9 (= 4.1 x 6.6/4.6). A Gompertz fertility model fitted to the average parities derived from such adjustments, shown in Figure 5b, suggests a total fertility rate for the district of 6.4 around 1988, the mid-point of the LEP-2 survey. Applying the model's age specific fertility rates to the adjusted population age distributions yields a general fertility rate of 165 per thousand women aged 15–49 and a crude birth rate of 48 per thousand.

Mortality estimates

The 112 886 individuals recorded at LEP-1 provided 407 000 person-years of follow-up in the 5-year period between LEP-1 and LEP-2, during which time 2 991 were reported to have died. It is not possible to estimate infant mortality in this population, since there are no data on deaths of children born in the

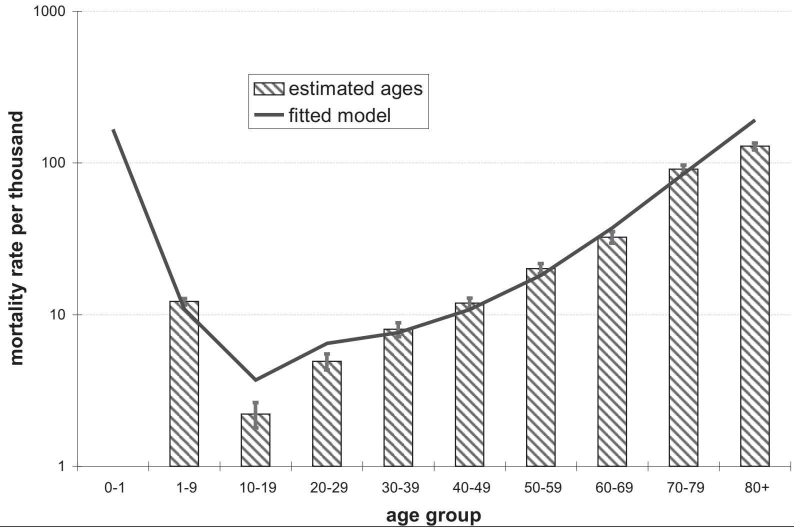
Figure 5b Mean CEB from adjustment of linked children per woman and fitted Gompertz relational fertility model



inter-survey interval, and those infants alive at LEP-1 would for the most part have already survived the dangerous neo-natal period, so that their subsequent mortality cannot be used to estimate mortality for the whole of the first year of life. For the same reason, we cannot directly calculate under-five mortality risks or conventional crude mortality rates, as infant deaths would be excluded. Observed age specific mortality rates are presented from age 1 upwards, but the values for ages 7 and under are based on the person-years lived by older children already born by LEP-1, and exclude the experience of those born in the inter-survey interval, although allowance has been made for left censoring.

Figure 6a shows the mortality rates obtained using information for the whole of the population with their estimated ages obtained after re-distributing those with unknown age as explained above. Figure 6b shows mortality data based only on those who were able to give a precise year of birth. The data are shown in ten-year age groups, so as to minimise the effect of inter-age group transfers following the age allocation procedure. Apart from the age-group 10–19, mortality rates were consistently higher in individuals who could only estimate their birth years using the local events calendar, compared to those reporting a precise year of birth. The age group 10–19 is

Figure 6a Observed and model fitted mortality rates for individuals with estimated years of birth: Karonga District, Malawi 1979–89

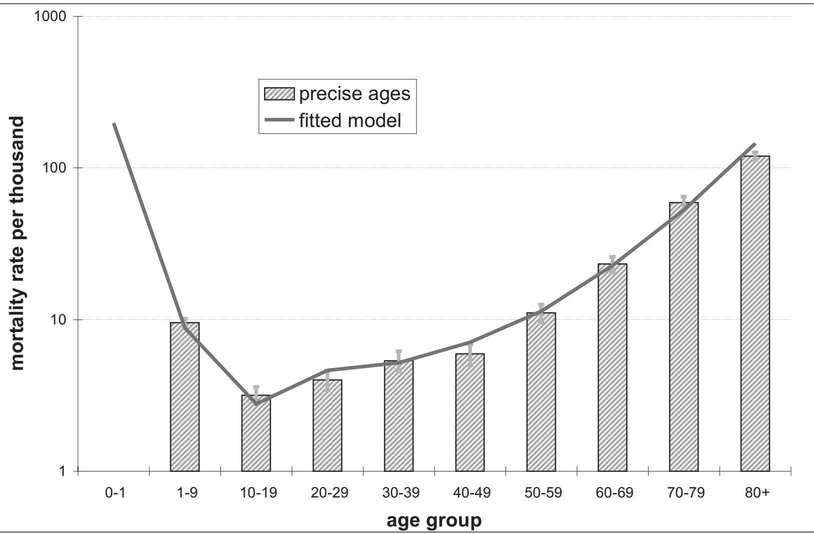


unusual in that more males than females have an unknown birth year (280 males compared to 80 females at LEP1).

Model life-tables (generated using the Brass logit relational system based on his general standard) were fitted to the observed data, by minimising the sums of squares of differences between observed and fitted logarithms of mortality rates in 10-year age groups. The within-age group structure can have an appreciable effect when relatively wide age intervals are used, and this was allowed for by assuming an average annual growth rate of 3 per cent during the birth years of persons enumerated in the two surveys, which was applied to the model person-years and events before comparing model and observed rates. As shown in Figures 6a and 6b, the models can be used to obtain extrapolated values for mortality rates in the youngest age group, and to reconstruct conventional mortality indices.

The model fitted to the whole dataset including the use of estimated ages, suggests an infant mortality rate of 123 per thousand, under-five mortality of 186 per thousand, an adult mortality index of 37 per cent (proportion of 15-year olds who would die before age 60), and life expectancy of 50 years. The model fitted only to the mortality experience of those with precisely stated ages implies infant mortality of 147 per thousand, under-five mortality

Figure 6b Observed and model fitted mortality rates for individuals with precise years of birth: Karonga District, Malawi 1979–89



of 198 per thousand, an adult mortality index of 26 per cent and life expectancy of 54 years. Fitting to the precise age data generates a model life-table with a somewhat extreme relationship between relatively low adult and very high child mortality, unlike the estimated age model fit, in which both adult and child mortality are high.

Indirect mortality estimation

The special surveys in Kasowa and Lower Songwe in 1993 and 1996 yielded information on proportion of children dead by mother's age that could be used to make indirect mortality estimates. Of 3 106 women interviewed in these surveys, 2 983 (96 per cent) indicated the total number of children they ever had. The proportion dead of CEB to women aged between 15–49 years increased with age of the women, from 23 per cent in those aged 15–19 up to 31 per cent in those aged 45–49 years. The calculation procedure for obtaining the indirect estimates of child mortality is shown in Table 1, using the Brass General standard model life-table.

Table 1 Brass Indirect Estimation of Child Mortality: Kasowa area, Karonga, Malawi 1993–96

Age of mother	*Average age x of children*	*Mean parity*	*Proportion dead*	*Indirect mortality q(x)*	*Indirect infant mortality rate q(1)*	*Indirect Under-5 mortality q(5)*	*Time location of mortality (year)*	*Indirect central mortality rate $_5M_0$*
15–19	1	0.3611	0.2268	0.1831	0.1831	0.3091	1992.6	0.074
20–24	2	1.5294	0.2105	0.2079	0.1517	0.2630	1991.3	0.062
25–29	3	3.0639	0.2380	0.2398	0.1548	0.2677	1989.5	0.065
30–34	5	4.1429	0.2348	0.2415	0.1376	0.2415	1987.4	0.060
35–39	10	5.7908	0.2563	0.2695	0.1321	0.2330	1985.1	0.063
40–44	15	5.8559	0.2887	0.2974	0.1397	0.2448	1982.6	0.068
45–49	20	6.5960	0.3093	0.3141	0.1348	0.2372	1979.6	0.069
Average for calendar years 1982–88					0.1365	0.2397		

Under-five mortality ranged between 240 and 270 per thousand and infant mortality from 135 to 155 per thousand in the period 2–15 years prior to the survey. The plot of child mortality suggests only slight fluctuations up to 1988 but between 1988 and 1993 there is a noticeable rise, even if the most recent point is ignored when interpreting the results of indirect analyses, as it is affected by the higher mortality of children of teenage mothers, and is not considered a reliable guide to the general trend (National Statistical Office Malawi and Macro International 1994, David, Bisharat, Hill *et al.* 1990).

Risk factor analysis for mortality

Socio-demographic risk factors for mortality are summarised in Table 2, which shows the observed mortality rates by sex, age, residence, education, occupation and housing characteristics. The crude rate ratios, comparing categories for each risk factor, are shown in the fourth column, and adjusted rate ratios (derived using Poisson regression) to allow for confounding by age and sex are shown in the sixth column. There was a relationship between mortality and geographical residence zone, with the highest mortality in the northern hills. Mortality decreased from the northern part down to the southern part of the district. Categorising zones into northern (zones A and B) versus southern Karonga (zones C, D and E), reveals that the individuals living in the northern part had 1.30 (95 per cent CI: 1.19, 1.39) times higher mortality than those in the southern part of Karonga District, after adjusting for age and sex.

Mortality was also found to be highest in those with no record of primary education (6.4 deaths per 1 000 PYR) and decreased with level of primary education, in casual labourers and those with no occupation (11.3 deaths per 1 000 PYR). Casual labourers had 2.30 (1.69, 3.15) times higher mortality than salaried workers and those with no primary education had 1.98 (1.59, 2.48) times higher mortality than those who had attended primary school and completed standards 6–8 after adjusting for age and sex. There were also differences in mortality based on housing quality (defined by construction materials) after adjusting for age and sex. Individuals living in houses made of wood and mud had 1.21 (1.10, 1.34) times higher mortality than those living in houses made of burnt bricks and those living in temporary shelters had 1.27 (0.99, 1.63) times higher mortality than those with burnt brick houses.

After adjusting for age, sex, occupation and housing, mortality was 1.14 (1.03, 1.25) times higher in individuals who could only estimate their birth years using the local events calendar compared to those with precise years of birth.

Table 2 Assessment of risk factors for mortality. Karonga District, northern Malawi 1979–89

Risk factor		*Crude estimates*			*Estimates adjusted for age and sex (derived using poisson regression)*	
		Number of deaths/ Person years (PYR)	*Crude mortality rate (per 1 000 PYR)*	*Rate ratio*	*Rate ratio (95 per cent CI)*	*P-value*
Sex	Female	1 255/187 530	6.7	1	1	
	Male	1 735/219 397	7.9	1.18	1.03 (0.91, 1.17)	0.744
Age (in years)	0–4	951/50 782	18.7	1	1	
	5–19	874/213 495	4.1	0.22	0.22 (0.20, 0.26)	<0.001
	20–44	482/104 184	4.6	0.25	0.26 (0.22, 0.31)	<0.001
	45+	683/38 466	17.8	0.95	0.65 (0.53, 0.79)	<0.001
Geographical zone	A Northern Hills	210/17 796	11.8	1	1	
	B North lake shore	1 261/155 766	8.1	0.69	0.68 (0.59, 0.79)	<0.001
	D Semi-urban	286/43 413	6.6	0.56	0.55 (0.46, 0.66)	<0.001
	C Southern Hills	723/106 500	6.8	0.58	0.56 (0.48, 0.66)	<0.001
	E South lake shore	511/83 547	6.1	0.52	0.49 (0.42, 0.58)	<0.001
Education	Std 6–8	342/78 886	4.3	1	1	
	Std 1–5	1 288/220 179	5.8	1.35	1.37 (1.20, 1.57)	<0.001
	No school (aged 5–18 years)	254/39 448	6.4	1.49	1.98 (1.59, 2.48)	<0.001

Occupation	Salaried worker	64/14 910	4.3	1	1	
	None & casual	105/9 301	11.3	2.63	2.3 (1.69, 3.15)	<0.001
	Fisherman/farmer	1 069/157 057	6.8	1.59	1.33 (1.03, 1.71)	0.031
	Traditional/trade	90/10 462	8.6	2.00	1.55 (1.12, 2.14)	0.008
Housing construction material	Burnt bricks	537/88 136	6.1	1	1	
	Wood & mud	1 522/204 244	7.5	1.23	1.21 (1.10, 1.34)	<0.001
	Sun-dried bricks/ pounded mud	662/85 779	7.7	1.26	1.28 (1.34, 1.43)	<0.001
	Grass temporary shelter	71/8 558	8.3	1.36	1.27 (0.99, 1.63)	0.060

Migration

Altogether 12 per cent (13 275) of individuals recorded at LEP-1 had out- migrated over the (approximate) 5-year period before LEP-2 (12 per cent for males, 11 per cent females). The rates of out-migration increased with age and peaked in the age group 20–24 (21 per cent for females, 24 per cent for males) and declined to very low levels for individuals aged over 45. The rates are higher for girls than boys under 15 years of age, but consistently higher for males above that age.

Of the 146 115 individuals recorded at LEP-2, 10 per cent (13 969) had in-migrated into Karonga District after LEP-1 (9 per cent males, 10 per cent females). In general, the percentages of in-migrants were higher for females than for males between the ages 10–29. This is likely to reflect the tradition in northern Malawi for women to join their husbands' households upon marriage.

DISCUSSION

This analysis of Karonga data provides a detailed description of a developing country population in sub-Saharan Africa during the 1980s. The dataset provides baseline information for further analyses of a variety of demographic, social, economic and epidemiological issues. Of particular importance today are questions relating to the demographic impact of HIV/AIDS, which has grown into the prime determinant of population change in the years since collection of the data analysed here.

The population of Karonga during the 1980s was predominantly young, typical of rural African communities, with 46 per cent of the population under 15 years of age. The age distribution reflects a high fertility population, consistent with our estimate of a total fertility of 6.4 children, and comparable to estimates of 5.7 and 6.7 for the northern region of Malawi derived from the 1987 census and 1992 MHDS respectively (National Statistical Office Malawi 1987, NSO Malawi and Macro International 1994).

The crude sex ratios (number of males per 100 females) at LEP-1 and LEP-2 were 90 and 94 with an excess of females over males for those aged between 20 and 55 years. This is partly attributable to migration of males to urban areas outside Karonga in search of better employment opportunities, and partly to lower (cumulated) female mortality. However after age 50 sex ratios increased with age, even after the adjustment for approximately reported age. This increase is unexpected because women live longer than men in most populations (Timæus 1997) so one would expect the sex ratios to continue to fall in older age groups. The observed trend may be in part a reflection of men returning to their homes after retirement, though this should not push up the sex ratio close to 1 because excess mortality would also affect the migrants. In a stable population with the kind of mortality patterns seen in this population sex ratios would be expected to be well below 0.9 in the oldest age group, 65+. It is hard to tell how much of the distortion is due to migratory movements. The increase in sex ratio at older ages may be due, in part, to errors in birth year reporting by females as females are less likely to give precise years and this increases with age. It is likely that birth year estimates are also affected by age exaggeration, and this may be worse for men than for women, since age confers status on those who aspire to leadership in the community.

Our estimate of crude birth rate (47 per thousand for the period 1982–88) is higher than the Karonga District figure of 37 births per 1 000 reported in the 1987 Malawi Census, but close to the DHS estimate for 1987–92 for the Northern region of 44 per thousand. The total fertility estimate of 6.4 births per woman is also close to the DHS Northern region estimate of 6.7, although the age-specific schedule in the DHS is older than that derived from our Gompertz model fit to mean parity data: the mean age of the DHS fertility schedule for 1987–92 is 29.8, whereas the mean age of our model is 26.4.

In adults, we observed higher mortality in individuals with birth estimates compared to those with precise years of birth. Knowledge of birth year is related to level of education and socio-economic status of individuals. The fact that more males than females are able to give a precise year of birth reflects the historical higher educational attainment of males, and a greater awareness of age in males, for whom advanced years confer greater social status. The links between social standing and age may lead to some age exaggeration among males purporting to know their true birth year. In the 10–19 age group more females than males report their precise year of birth – this may reflect a greater awareness of age among teenage females, for whom puberty signifies availability for marriage. Again, some exaggeration of age may occur for girls who experience puberty early, or begin childbearing at an atypically early age. Census reports are also subject to age misreporting, but the longitudinal nature of the LEP-1 and LEP-2 data, and the special effort made to distinguish known date of birth from approximations provides more detailed insights into the likely patterns of miss-statement.

Geographical patterns of mortality may reflect a variety of environmental factors. North Karonga is hilly in parts, but includes the flood plain of the Songwe River, which renders large parts of the area inaccessible during the rainy season, and exacerbates malaria transmission. The southern part of the district consists of a small urban centre and a less hilly area with easier access to the lake and town (to supplement food and income), and to health centres. There was a higher leprosy incidence in the north than south Karonga (Sterne, Ponninghaus, Fine *et al*. 1995). Recently, it has been found that filariasis is also higher in the north (Ngwira, Jabu, Kanyongoloka *et al*. 2002). These geographical patterns deserve further investigation.

Since there are no data on births occurring between LEP-1 and LEP-2, it is not possible to make direct estimates of infant and child mortality based on children surviving to LEP-2. Nor is it possible to generate a complete life-table for the population or to estimate the crude death rate. However, two types of indirect estimate are possible: one based on model life-table fitting to conditional survival estimates for the population aged 5 and over, the second based on Brass procedures applied to data on child survival collected in selected villages.

If we restrict the model life-table fitting to mortality data based on individuals for whom a precise date of birth is known, the survival estimates for older adults are based on an increasingly selected low-mortality population. This increasing selectivity with age affects females more than males. Model life-table survival curves generated from these data would be expected to have too shallow a slope, leading to over-estimates of infant mortality. Indeed, the DHS estimate for the northern region for the period 1982–87 of 123 per thousand, matches exactly the estimate obtained for the fit to the full dataset with estimated ages, but is considerably lower than the estimate of 147 based on fitting only the "precise" age mortality data. Under-five mortality is less affected by this bias, so that both our estimates: 186 based on the total population fitting, and 198 based on the restricted fitting to those with precise age information, are close to the DHS estimate of 202 per thousand.

A crude death rate (CDR) estimate of 18.3 was obtained by applying the model life-table obtained by fitting to the experience of the total population (including those with estimated date of birth) to the adjusted population age distribution, and the equivalent measure based only on those with precise date of birth was 16.5. There was very little difference between estimates based on LEP-1 or LEP-2 populations. By comparison, the CDR for Karonga District from the 1987 Malawi censuses was 17.4 deaths per 1 000 persons, based on household deaths in the last 12 months, corrected for under-enumeration (National Statistical Office Malawi 1987).

The retrospective estimates from the Brass indirect procedure which is restricted to selected villages in the north of the district yield average values of 137 and 240 for infant and under-five mortality respectively over the same time period (1982–87) corresponding to the middle of the LEP-1 to LEP-2 interval. These villages are thought to constitute a relatively high mortality area within the northern part of Karonga district (which itself is characterised by higher mortality rates than the southern part of the district), so the relationship of these mortality risks to the DHS estimates, and the model life-table extrapolation for under-five mortality is plausible.

There is a suggestion from the indirect analysis that under-five and infant mortality began to rise after 1988, and it is tempting to ascribe this to the HIV epidemic. Most infants who acquire HIV from their mothers die in the first five years of life, and the estimated proportion of infected births in Karonga district rose slowly from about 1.5 per cent in 1988 to 4.5 per cent in 1992 (Crampin *et al*. 2003). However, the method of indirect estimation of child mortality does not take into account the complex relationship between HIV/AIDS, maternal age, fertility and maternal survival, and would tend to smooth out any real trends due to HIV infection. In fact, national level data from Malawi suggest that improvements in non HIV-related child mortality have been the dominant feature governing child mortality trends since 1990 (Zaba, Marston and Floyd 2004). The apparent trend in the more recent Brass estimates is most likely an artefact due to the higher mortality of children born to teenage mothers (David, Bisharat, Hill *et al*. 1990).

HIV prevalence measured for subpopulations in this area on the basis of retrospectively testing adult blood samples collected during LEP-1 and LEP-2 are 0.2 per cent and 4 per cent for 1982 and 1988 respectively, but more recent estimates based on ante-natal surveillance imply that HIV prevalence in the district is currently around 13 per cent (Crampin *et al*. 2003). The patterns of fertility, mortality and age/sex structures presented above relate largely to the time before the HIV epidemic took hold in this area, and are likely to change appreciably as a consequence of high AIDS-related mortality. The Karonga setting provides an exceptional opportunity to investigate the demographic effects of the HIV epidemic on this rural population. An HIV cohort study following up HIV-positive individuals, HIV-negative controls and their families, has been implemented, and showed that between 1988 and 1998 mortality among those who were infected by 1988 was more than eight times as high as mortality among the uninfected, mortality amongst the spouses of HIV-infected individuals was four times as high as among the spouses of HIV-negative controls, and children of HIV-infected mothers had almost three times the mortality rate of children of uninfected mothers (Crampin, Floyd, Glynn *et al*. 2002, 2003). A demographic surveillance system has now been implemented – affiliated to the INDEPTH network (INDEPTH 2004) – to track fertility and mortality trends. The results of these new studies will be compared with the baseline measures presented in this paper.

Acknowledgement

The LEP was funded by the British Leprosy Relief Association (LEPRA) with contributions from the British Overseas Development Administration and the WHO/UNDP/World Bank Special Programme for Research and Training in Tropical Diseases (WHO/TDR). The KPS is currently funded by LEPRA and the Wellcome Trust. Tobias Chirwa was supported by WHO/TDR during his work at London School of Hygiene and Tropical Medicine. We thank Lyn Bliss, Keith Branson and the many field and data management staff who collected these data over the past 24 years, and the Health Sciences Research Committee of the Malawi Ministry of Health and Population for encouraging and approving the work.

References

Crampin A.C., Floyd S., Glynn J.R., Sibande F., Mulawa D., Nyondo A., Broadbent P., Bliss L., Ngwira B. and Fine P.E.M. 2002. "Long term follow-up of HIV positive and negative individuals in rural Malawi", *AIDS* 16:1545–50.

Crampin A.C., Floyd S., Glynn J.R., Madise N., Nyondo A., Khondowe M.M., Njoka C.L., Kanyongoloka H., Ngwira B., Zaba B. and Fine P.E. 2003. "The long-term impact of HIV and orphanhood on the mortality and physical well-being of children in rural Malawi", *AIDS* **17**(3):389–97.

Crampin A.C., Glynn J.R., Ngwira B.M.M., Mwaungulu F.D., Ponnighaus J.P., Warndorff D.K. and Fine P.E.M. 2003. "Trends and measurement of HIV prevalence in northern Malawi, 1988–2001", *AIDS* 17:1817–25.

David P.H., Bisharat L., Hill A.G. and Bennett S. 1990. *Measuring childhood mortality: a guide for simple surveys.* Amman, Jordan: UNICEF.

Glynn J.R., Ponnighaus J., Crampin A.C., Sibande F., Sichali L., Nkhosa P., Broadbent P. and Fine P.E. 2001. "The development of the HIV epidemic in Karonga District, Malawi", *AIDS* **15**(15):2025–9.

Hill K. 1991. "Approaches to the measurement of childhood mortality: A comparative review". *Population Index* **57**(3):368–82.

INDEPTH network. 2002. "Population, Health and Survival at INDEPTH Sites", Ottawa, Canada: IDRC. ISBN: 0-88936-948-8.

INDEPTH. 2004. http://www.indepth-network.org/

KPS website. 2004. http://www.lshtm.ac.uk/ideu/kps/

Lesthaeghe R. and Jolly C. 1992. "The start of the sub-Saharan fertility transition: some questions and many answers", IPD-Working Paper 1992–4, Brussels Belgium, Inter-university Programme in Demography, Centrum Sociologie, Vrije Universiteit, Brussels.

National Statistical Office Malawi. 1987. *Malawi Population and Housing Census: Population Characteristics.*

National Statistical Office Malawi and Macro International, 1994. *Malawi Demographic and Health Survey, 1992*. Calverton, Maryland: Macro International.

Ngwira B.M.M., Jabu C.H., Kanyongoloka H., Mponda M., Crampin A.C., Branson K., Alexander N.D.E. and Fine P.E.M. 2002. "Lymphatic filariasis in Karonga District, N Malawi: A prevalence survey", *An of Trop Med & Parasitology* 96:137–44.

Ponnighaus J.M., Fine P.E., Bliss L., Gruer P.J., Mwamondwe B.K., Msosa E., Rees R.J., Clayton D., Pike M.C., Sterne J.A., *et al.* 1993. "The Karonga Prevention Trial: a leprosy and tuberculosis vaccine trial in northern Malawi. I. Methods of the vaccination phase", *Lepr Rev* **64**(4):338–56.

Ponninghaus J.M., Fine P.E., Bliss L., Sliney I.J., Bradley D.J. and Rees R.J. 1987. "The Lepra Evaluation Project (LEP), an epidemiological study of leprosy in Northern Malawi. I. Methods", *Lepr Rev* **58**(4):359–75.

StataCorp, Stata Statistical Software: Release 7.0. 2001. Stata Corporation: College Station, TX.

Sterne J.A.C., Ponnighaus J.M., Fine P.E.M. and Malema M. 1995. "Geographic determinants of leprosy in Karonga District, Northern Malawi", *International Journal of Epidemiology* 24:1211–22.

Timæus, Ian. 1997. "Mortality in Sub-Saharan Africa", in *Symposium on Health and Mortality*.

Zaba B., Marston M., Floyd S. 2004. "The effect of HIV on Child Mortality Trends in sub-Saharan Africa", Accepted for publication in *UN Population Bulletin*. http://www.un.org/esa/population/publications/adultmort/Adultmortality.htm

APPENDIX A

Age redistribution

Age heaping is often observed due to digit preference in the face of uncertainty about actual birth dates. When birth date reporting is poor, local events calendars are often used to give approximate years of birth. Although in this analysis age heaping certainly dominates the picture due to the local events calendar, it would be unusual not to have some age misreporting due to digit preference as well. The procedure presented here for smoothing the age distribution by redistributing age of individuals with birth estimates based on those who gave precise years of birth may be applicable to other studies.

The age distribution in Figure 2a showed apparent age heaping in the older age groups, which was a reflection of birth-year estimates used. The age distribution of the population is skewed and, as such, the middle of an estimated period of birth may not be a good representation of the birth estimate of individuals born within that time period.

Individuals with birth year estimated according to the local events calendar were counted and classified according to the birth-year range. Individuals with precise years of birth were also counted according to the corresponding ranges of the event calendar, yielding an observed birth-year distribution of individuals who were born within each estimate period. Those with birth-year estimates were then redistributed proportionally according to the age distribution of those with precise ages within the same calendar range

As a result of this smoothing, the age distribution in Figure 2b does not show any digit preference. There is a smooth decline with age except for a hump at about 70 years, which may reflect an artifact (i.e. old people claiming to be precisely 70 years of age, and field staff then subtracting 70 from year of interview, and recording that figure as a precise birth year).

In our age redistribution by proportional allocation we have assumed that individuals in a particular age range estimate are a random sample of that birth cohort. This, of course, may not be true – as even within an age range, the proportion with precise years of birth will decline with age. It would be possible to model and correct for this trend. The smoothing methodology explained above would be appropriate in other populations with long time periods between consecutive events in a local reference calendar.

The estimation of infant mortality from proportions dying among births in the past 24 months

John Blacker[1] and William Brass[2]

This paper incorporates the last contribution to demographic methods of the late Professor William Brass before he died on 11 November 1999. An earlier version was published in 1999 in the series of Research Papers of the Centre for Population Studies at the London School of Hygiene and Tropical Medicine (Brass and Blacker 1999). This version has been extensively revised and expanded.

Abstract
Censuses and demographic surveys in developing countries frequently include questions of adult women as to the date of their most recent live birth, and whether or not the child is still alive. Using a mathematical model of mortality in infancy and childhood, we show that the proportion dying among children born in the 24 months prior to the census or survey, multiplied by a factor of 1.09, gives a close approximation to the infant mortality rate. Applications to data from Indonesia and Kenya give results which are in good agreement with those from other sources. The method has the advantages that it provides up-to-date estimates which are robust to the selectivity bias inherent in birth history data for countries with a high prevalence of HIV. But it is vulnerable to reporting errors and non-response.

Key words
Infant mortality rate; births in the last 24 months; models of mortality in infancy and childhood; HIV selectivity bias

INTRODUCTION

Most African censuses and demographic surveys have included questions of adult women as to the number of children they have borne, how many are still living and how many have died. From the proportions of children dying,

1 Centre for Population Studies, London School of Hygiene and Tropical Medicine
2 Formerly at the Centre for Population Studies, London School of Hygiene and Tropical Medicine

tabulated by age groups of mother, indirect estimates of infant and child mortality can be obtained, using the techniques originally devised by Brass (1968) and the subsequent variations on this theme (United Nations 1983).

Extensive applications of these methods have revealed various problems. In particular the estimates based on the youngest age group of mothers, 15–19, have generally shown a sharp upward bias which has been attributed to their large component of first births which are generally subject to above-average mortality (Ewbank 1982). This bias is also sometimes visible, though to a lesser extent, in the data for the 20–24 age group. Thus the data for women aged under 25 frequently have to be discarded, and the first useable estimates are those for the 25–29 age group. But these estimates typically have a time location about five years prior to the census or survey, so that we have no information on recent trends in mortality.

This difficulty has now been further compounded, in countries with high prevalence of HIV and AIDS mortality, by the selectivity of the respondents. We only have reports from living mothers; the mortality of children borne by mothers who have died will not feature in our data. If, as must surely be inevitable in any circumstances, there is an association between the mortality of the mothers and that of their children, the estimates will be biased downwards. In normal conditions such bias is trivial, since female mortality at these ages is relatively low, but the advent of the AIDS epidemic has rendered it far from negligible. Not only has the mortality of women in the child-bearing years enormously increased, but the association with that of their children has been greatly strengthened by the vertical mother-to-child transmission of the virus. It will clearly have affected not only the indirect mortality estimates from census data, but also direct estimates from birth histories, such as are collected in the Demographic and Health Surveys.

Much research remains to be done on the measurement of this selectivity bias, its relation to the levels and trends of HIV prevalence, and possible methods of correction. However some exploratory modelling by Basia Zaba and Marc Artzrouni has suggested that the bias is relatively small for children born less than five years before the census or survey, but becomes progressively greater beyond that point (Mahy and Zaba 2005; Artzrouni and Zaba 2003). The implications of this finding for the traditional indirect estimates are clearly serious: we cannot use those based on the reports of females under 25 because of the first-birth bias; nor can we trust those of mothers over 25 because of the selectivity. Demographers therefore need to look for alternative sources of information on mortality which will give estimates which are both up-to-date and robust to selectivity. This paper suggests one such possible source.

In many censuses and surveys questions have been asked of adult women as to the date of their most recent live birth and whether or not the child is still alive. Estimates of the infant mortality rate (IMR) have then been made by calculating the proportion dying among children born during the last 12 months, which is assumed to represent about two-thirds of the IMR. The results of this procedure have varied: sometimes they have given estimates which are plausible and consistent with those derived from other information, but all too often they have clearly been unacceptable (Blacker 1984; Chackiel and Gough 1989). The reasons are not hard to find. Census questions on date of last birth are vulnerable to non-response; some enumerators do not bother to ask the question, and enter an answer based on the age of the youngest child in the household; many respondents give the date of birth as "about a year ago", so there is heavy heaping on the date 12 months before the census or survey.

In these circumstances there are obvious advantages in basing the calculations, not on births in the last 12 months, but on those occurring during the *24* months prior to the survey. The number of births would be approximately doubled, thus reducing the sampling errors, and the problem of how to handle the births heaped on a date just 12 months before the survey would be largely eliminated. Only a very small proportion of the mothers will have had more than one birth in successive pregnancies during the period, so that any bias resulting from the under-representation of children borne by mothers with short birth intervals would be minimal. The data should also be robust to the AIDS-selectivity bias. As mentioned above Artzrouni and Zaba found that this bias was small for children born in the last five years; *a fortiori* it will be even smaller for those born in the last two years; and on common-sense grounds we know that the period of HIV incubation is much longer for adults than for children, so the proportion of HIV-positive women who were fit enough to give birth in the last two years who will die before the census or survey will be very small. Furthermore the proportion dying will clearly be much closer to the infant mortality rate: while the average exposure of children born in the last 12 months is about half a year, that of children born between 12 and 24 months ago will be about 1.5 years, so the proportion dying of the two combined should be a close approximation to the IMR. However this relationship needs to be examined more closely.

Using standard life-table notation, the proportion surviving from births in the last 12 months is ${}_1L_0$, and that from the 12 months before that is ${}_1L_1$. Thus the proportion dying among the births of the last 24 months will be

$$1 - \frac{{}_2L_0}{2},$$

when the life-table survivors, *l(x)*, and the life-table population, ${}_nL_x$, are calculated from a radix of 1. The relationship between this and the infant mortality, which we will denote at $q(1)$, will vary from population to population, depending on the age pattern of mortality in infancy and childhood. In model life-table systems, such as those of Coale and Demeny, with discreet age patterns and fixed separation factors for the calculation of ${}_1L_0$, it is a fairly simple matter to derive the relationship between $1–{}_2L_0/2$ and $q(1)$. However it would clearly be preferable if we could use a continuous function which would cover the whole range of human experience, rather than discreet models.

A MATHEMATICAL MODEL OF MORTALITY IN INFANCY AND CHILDHOOD

Over the first few years of life mortality initially falls rapidly with age and then levels out. It clearly resembles a hyperbola, $y = 1/x$, and various demographers have used this relationship to depict early-age mortality (e.g. Keyfitz 1966). In accordance with this line of thought, we suggest that the curve of life-table survivors *l(x),* can be quite well represented by the simple function:

$$l(x) = (1 + \alpha x)^{-\beta}$$

where x is age in either years or months; alpha and beta are two constants which give the level and pattern of the death rates[3]. Their nature is most conveniently seen in terms of $\mu(x)$, the force of mortality at age x. Taking logs and differentiating:

$$\ln l(x) = -\beta \ln(1+\alpha x), \text{and } \mu(x) = \frac{\alpha\beta}{(1+\alpha x)} = \frac{\mu(o)}{1-\alpha x}.$$

Alpha thus determines the pace of fall in mortality with age, and beta adjusts for the level.

3 The values of the parameters will of course differ according to whether months or years are used, but the end result, in terms of the fitted life-table values, will be the same. The use of months is clearly preferable if one needs to fit the model to obtain neonatal rates or other subdivisions of the first year of life, as we have done with the Matlab and Sine Saloum data shown below. Otherwise years are generally simpler to use, as we have done with the INDEPTH life-tables shown in Table 3 below.

The model may be fitted to any life-table, given two values of $l(x)$, say at ages 1 and 5 years, or 12 and 60 months. Alpha can then be derived from the ratio of their logarithms, which we denote as R. Then

$$R = \frac{\ln l(12)}{\ln l(60)} = \frac{\ln(1+12\alpha)}{\ln(1+60\alpha)}.$$

The value of alpha which gives the observed value of R can then be found by iteration, which, with a spreadsheet or programmable calculator, can be done very quickly. Beta can then be determined as

$$\frac{-\ln l(60)}{\ln(1+60\alpha)}.$$

The choice of the two ages at which to fit the model can be found empirically. Other more refined methods of fitting, such as minimising the squared differences between the observed and fitted values, can of course be used, but have been found to differ only minimally from this simpler procedure.

By way of illustration we have fitted the model to two data sets with widely different levels and patterns of mortality. The first is the 1992 data from Matlab in Bangladesh, where the population has been kept under observation by the International Centre for Diarrhoeal Disease Research (ICDDR, B 1995), and where mortality between 1 and 5 years is relatively low in comparison with that in infancy. The model was fitted at 6 and 60 months. The results are shown in Table 1.

Table 1 Matlab 1992: Observed and model values of life-table survivors in the first ten years of life

Age (months) x	*Observed 1(x)*	*Fitted 1(x)*
0	1 000	1 000
1	948	956
6	925	925
12	914	913
24	900	900
36	891	893
48	887	888
60	884	884
120	877	872
Ln *l*(6) = –0.078178	Ln *l*(60) = –0.122914	R =0.636038
	Alpha = 8.8845	Beta = 0.0196

The largest discrepancy between the observed and fitted values is, unsurprisingly, in the first month of life. The fits at 12, 24, 36 and 48 months are highly satisfactory; even the extrapolated figure at age 10 years is reasonably close to the observed value.

The second fitting is to the data from Sine Saloum, Senegal, in the 1960s, where almost half the children died in the first five years of life, and where the mortality between 1 and 5 was substantially higher than that in infancy (Cantrelle and Leridon 1971). This provides a more exacting test, and in one respect the model cannot hope to reproduce the original data. The observed age-specific mortality rates actually increased in the second six months of life, and did not start to fall again until well into the second year; in contrast the model can only feature a monotonic decline in mortality with increasing age. Nevertheless by fitting the model at 12 and 60 months, it reproduces the high childhood mortality in comparison with infancy, as shown in Table 2. Thus the flexibility of the model and its ability essentially to represent the range of human experience is demonstrated.

Table 2 Sine Saloum 1962–68: Observed and model life-table survivors in the first five years of life

Age (months) x	*Observed 1(x)*	*Fitted 1(x)*
0	1 000	1 000
3	920	931
6	886	875
9	842	829
12	789	789
24	629	676
36	548	602
48	521	549
60	509	509
Ln l(6) = −0.23638	Ln l(60) = −0.67565	R = 0.349856
	Alpha = 0.056479	Beta = 0.456818

MODELS OF MORTALITY AMONG CHILDREN BORN IN THE LAST 24 MONTHS

The proportion dead, denoted by D, among children born in the last 24 months, is, as we have seen,

$1 - \frac{{}_2L_0}{2}$; it may also be written as $1 - \frac{1}{2}\int_0^2 lxdx$.

Inserting the function for *lx* described above

$$\int_0^2 lxdx = \frac{[(1+2\alpha)^{1-\beta} - 1]}{\alpha(1-\beta)}.$$

The infant mortality rate, $q(1)$, is $1-(1+\alpha)^{-\beta}$. Thus the adjustment factor needed to convert D into $q(1)$, can readily be calculated given values of α and β.

Using a spreadsheet, we have generated 120 models with alpha ranging from 0.2 to 1000 and beta from 0.01 to 0.8, and so with correspondingly wide ranges in the levels and patterns of mortality: infant mortality, $q(1)$, from 26 to 478 per thousand, and the proportion which infant mortality forms of deaths in the first two years of life, $q(1)/q(2)$, from 57 to 92 per cent. We also calculated the adjustment factor, $q(1)/D$, needed to convert the proportion dead among children born in the last 24 months into infant mortality. Examination of the results shows that $q(1)/D$ is unrelated to the level parameter, β, and is essentially determined by the age pattern, as defined by α. The nature of the relationship, which is perhaps somewhat surprising, is shown in Figure 1, where $q(1)/D$ is plotted, not against α, but against the more readily understandable $q(1)/q(2)$. In the central part of the range, with $q(1)/q(2)$s of between 65 and 85 per cent, the correction factors nearly all lay between 1.08 and 1.1, with a mean of 1.092. At both ends of the distribution they drop away, but do not fall below 1.04, at least within the range we have created.

Some of the combinations of α and β in this array of models may be regarded as somewhat unrealistic. We have therefore fitted our model to some real African life-tables and so calculated the values of $q(1)/D$. The INDEPTH publication *Population and Health in Developing Countries* (2002) contains abridged life-tables, by sex, for 17 African demographic surveillance sites. They cover a broad range of infant and child mortality. Under-five mortality ranges from 32 per thousand for males in Agincourt (South Africa) to 255 per thousand for males in Bandafassi (Senegal). The parameters of the models were fitted to the life-table survivors at ages 1 and 5 years. The results are shown in Table 3. It will be seen that, of our 34 life-tables, 24 had values of $q(1)/D$ of between 1.087 and 1.097. The median value was 1.092.

We have also made these calculations for two populations with low mortality – England and Wales (1993–95) and Japan (2001). These gave slightly lower estimates of $q(1)/D$ – 1.04 for England and Wales and 1.07 for Japan. In such low mortality populations, relatively greater reductions have been achieved in child than in infant mortality, so that the $q(1)/q(2)$ and the α tend to be high.

We may conclude therefore that estimates of infant mortality derived from proportions of children dying among those born during the last 24 months are remarkably robust to variations in the level and age pattern of infant and child mortality. If such estimates are being made for a population for whom nothing is known about the age pattern, a "default" correction factor of 1.09 can be used. If, however, there is reason to believe that child mortality is either unusually high in relation to that in infancy, as in parts of West Africa, or unusually low, such as would merit the use of the Coale–Demeny "East" or the United Nations "Chilean" models, then a slightly lower figure, not below 1.04, may be used. The use of lower factors might also be appropriate where the level of mortality is low – say infant mortality rates of under 50 per thousand. In any event the error introduced by the use of an inappropriate correction factor will be almost certainly be trivial compared with those due to response errors in the data.

Figure 1 Children born in the last 24 months – estimation factors for infant mortality

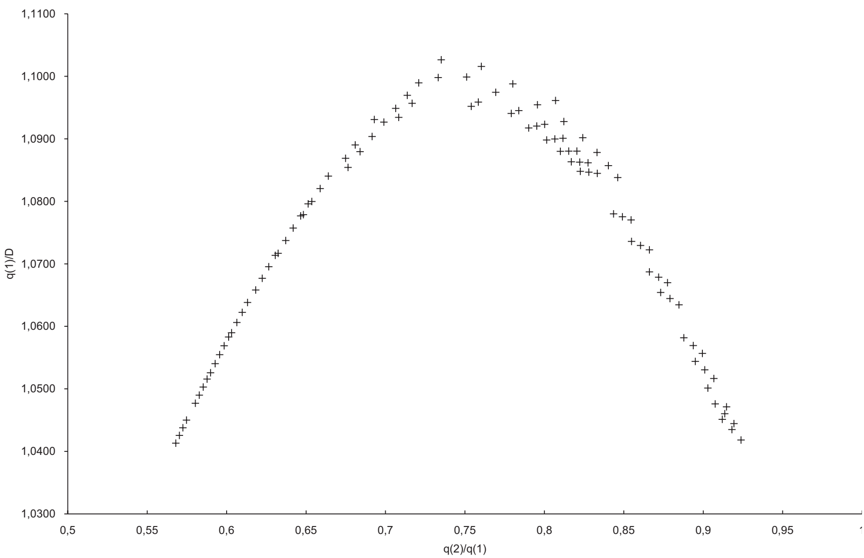


Table 3 Mortality models and estimates of $q(1)/D$ for African surveillance sites

Surveillance Site	*Date*	*Sex*	*1(1)*	*1(5)*	*Alpha*	*Beta*	*q(1)/D*
Agincourt	1995–99	M	0.98494	0.96768	2.074	0.0135	1.0872
South Africa		F	0.98337	0.96631	2.700	0.0128	1.0909
Bandafassi	1995–99	M	0.86140	0.74546	3.245	0.1032	1.0967
Senegal		F	0.88357	0.78258	3.150	0.0870	1.0958
Bandim	1995–96	M	0.88763	0.77243	2.068	0.1063	1.0919
Guinea-Bissau		F	0.89848	0.78320	1.650	0.1098	1.0880
Butajira	1995–97	M	0.93438	0.88044	4.185	0.0412	1.0953
Ethiopia		F	0.92891	0.87113	4.250	0.0445	1.0955
Dar es Salaam	1994–99	M	0.93362	0.88614	6.105	0.0350	1.0951
Tanzania		F	0.93280	0.88384	5.790	0.0363	1.0952
Farafenni	1995–99	M	0.93196	0.82900	0.893	0.1105	1.0723
Gambia		F	0.93354	0.83168	0.869	0.1100	1.0715
Gwembe	1991–95	M	0.89433	0.78826	2.229	0.0953	1.0925
Zambia		F	0.90330	0.81108	2.608	0.0793	1.0937
Hai	1994–99	M	0.93322	0.90827	57.500	0.0170	1.0716
Tanzania		F	0.94346	0.91829	29.150	0.0171	1.0801
Ifakara	1997–99	M	0.92388	0.87563	8.440	0.0353	1.0934
Tanzania		F	0.91391	0.86797	14.100	0.0332	1.0887
Manhica	1998–99	M	0.91425	0.85125	5.375	0.0484	1.0958
Mozambique		F	0.94063	0.88381	2.871	0.0452	1.0931
Mlomp	1995–99	M	0.95176	0.91120	4.130	0.0302	1.0949
Senegal		F	0.95058	0.90140	2.672	0.0390	1.0921
Morogoro	1994–99	M	0.89476	0.81668	4.955	0.0623	1.0963
Tanzania		F	0.88806	0.81810	7.960	0.0541	1.0942
Navrongo	1995–99	M	0.89342	0.81879	5.805	0.0588	1.0959
Ghana		F	0.89704	0.83135	7.690	0.0503	1.0944
Niakhar	1995–98	M	0.91020	0.77655	0.860	0.1516	1.0732
Senegal		F	0.92784	0.80802	0.689	0.1429	1.0659
Nouna	1995–98	M	0.96569	0.86185	0.099	0.3697	1.0197
Burkina Faso		F	0.95729	0.85503	0.266	0.1854	1.0381
Oubritenga	1995–98	M	0.89775	0.81159	3.542	0.0713	1.0959
Burkina Faso		F	0.90812	0.81291	2.142	0.0842	1.0913
Rufiji	1999	M	0.85246	0.82045	817.0	0.0238	1.0454
Tanzania		F	0.82440	0.79712	10 250.0	0.0209	1.0322

APPLICATION TO DATA FROM INDONESIA

In late 1992 and early 1993 a survey of infant and child mortality in three provinces of Indonesia was conducted by the Ministry of Health of the Government of Indonesia, with support from the World Health Organization and the London School of Hygiene and Tropical Medicine (Blacker, n.d.). Indonesia is a country with unusually large geographical differentials in infant and child mortality. Indirect estimates of $q(5)$ for the 28 provinces from the 1990 census ranged from under 50 to over 200 per thousand live births. The survey aimed to investigate factors underlying these differentials, principally in terms of cause of death. Three provinces were therefore selected: one with very high mortality, Nusa Tenggara Barat; one with intermediate mortality, Sumatera Selatan; and one with low mortality, Yogyakarta. A cluster sample of approximately 3 000 ever-married women was drawn from each province.

The survey did not include full birth histories, but the ever-married women were asked for the dates of their last and next-to-last births, and whether or not the children were still alive. Women whose last births had occurred during the 24 months before the survey were extracted and tabulated by the survival of the children. The "default" conversion factors of 1.09 were used for Nusa Tenggara Barat and Sumatera Selatan, but for Yogyakarta it was reduced to 1.06. The results are shown in Table 4.

Table 4 Estimates of infant mortality from proportions dead among children born in the last 24 months for three provinces of Indonesia, 1992

	Births in the last 24 months	*Number of children dead*	*Proportion dead*	*Conversion factor*	*Estimated IMR*
Nusa Tenggara Barat	833	83	0.0996	1.09	109
Sumatera Selatan	793	45	0.0567	1.09	62
Yogyakarta	432	16	0.037	1.06	39

The evaluation of these results is problematical. The 1992 survey also asked questions on children ever born, living and dead, from which indirect estimates of mortality have been obtained. Similar estimates have also been derived from the 1980 and 1990 censuses. However the conversion of the estimates of $q(2)$, $q(3)$, $q(5)$, etc., into $q(1)$ is sensitive to the choice of model life-tables. Such estimates also tend to stop short some two or three years before the time point represented by the deaths among children born in the last two years. However the comparisons, for what they are worth, are shown in Figure 2. The indirect estimates were obtained using the United Nations computer package QFIVE, with the "South" models for Nusa Tenggara Barat and Sumatera Selatan and the "West" for Yogyakarta. They give the impression that the rates derived from the deaths among children born in the last two years were too low in Nusa Tenggara Barat and Sumatera Selatan, and too high in Yogyakarta. However the sampling errors attached to the estimates based on the recent births were substantial, and the divergence of these points from lines extrapolated from the indirect estimates would lie well within the 95% confidence intervals, which are also indicated on the graphs.[4]

APPLICATION TO THE 1999 KENYA CENSUS

The indirect estimates of child mortality derived from the 1962, 1969, 1979 and 1989 censuses of Kenya presented an unusually consistent series showing a steady downward trend up to the 1980s. But with the 1999 census problems were encountered. Apart from the biases described in the introduction, additional errors were introduced during the data processing, largely arising from the use of optical character readers for the "data capture". These machines frequently mis-read the numbers, and the incidence of these errors varied considerably between different parts of the census form. The field most severely affected was the column for dead female children, so that the proportions dead, and hence the mortality estimates, were seriously inflated. Heavy and judicious editing succeeded in reducing the effect of the errors, but the margin of uncertainty was inevitably enlarged. Indeed the untrustworthiness of the estimates continued to be reflected in the fact that the mortality rates for female children appeared to be higher than those for males – a finding which was at variance with all the other data sources on Kenya mortality. Nevertheless the results also strongly suggested that the downward trend in infant and child mortality had come to an end.

This conclusion is supported by the direct mortality estimates from the four Demographic and Health Surveys conducted in 1989, 1993, 1998 and 2003. Though less consistent than the indirect census estimates, they suggest a generally rising trend, as shown in Figure 3.

4 The estimation of confidence intervals is described in Appendix Three of Blacker (n.d.), the more significant portions of which are reproduced here in the Appendix.

Figure 2 Estimates of infant mortality, Indonesia

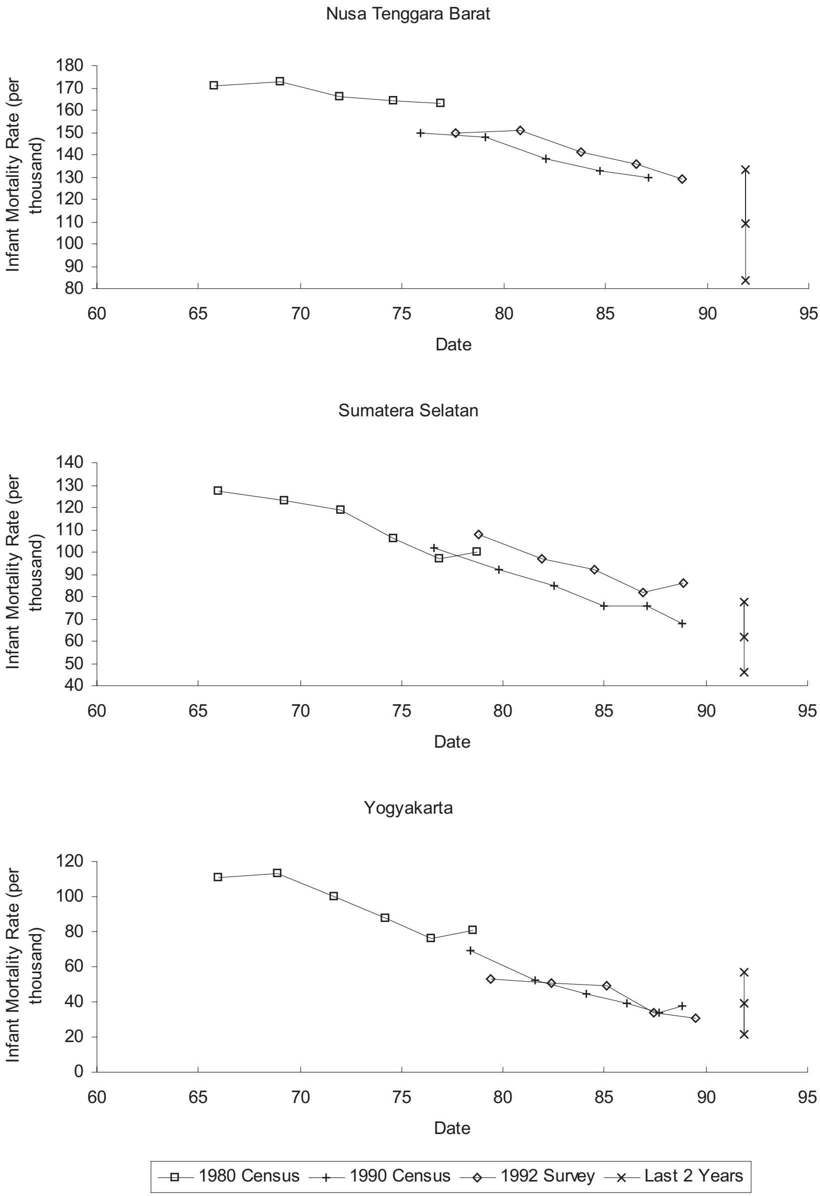
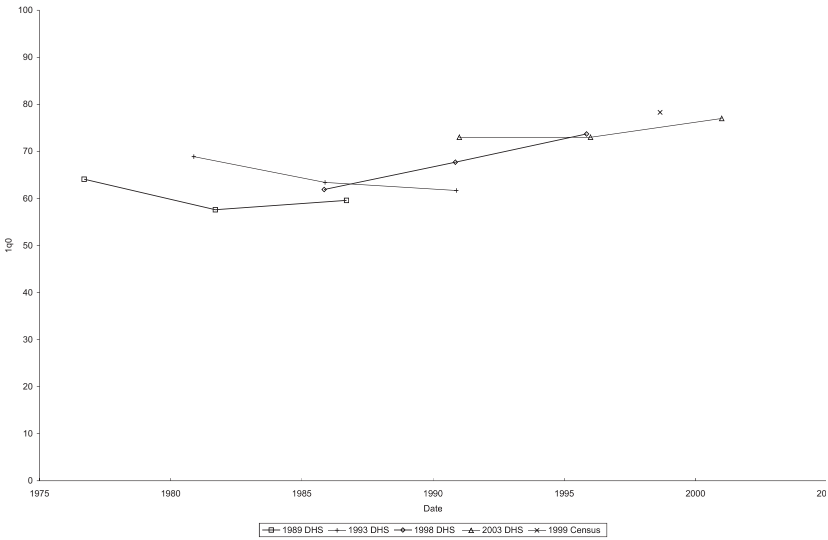


Figure 3 Kenya infant and child mortality – estimates from DHSs and the 1999 census



It should be noted, however, that these direct estimates from the DHSs were subject to quite large sampling errors: for example the estimated IMR of 77 per thousand for the five years prior to the 2003 survey had a relative standard error of 7.2 per cent, so that the 95% confidence limits of the estimate ranged from 66 to 88 per thousand (CBS/MOH/ORC Macro 2004, table B.2, p.267).

In this fog of uncertainty the estimates derived from the 1999 census data on the mortality of last-born children were a welcome addition. They are shown in Table 5 and the point for both sexes has been inserted in Figure 3 for comparison with the DHS estimates. Being based on 100% census enumeration comprising some 1.7 million births in the preceding 24 months, the question of sampling errors does not arise. It may also be noted that, in contrast to the indirect estimates from children ever born, the sex differentials are in the right direction.

Table 5 Estimates of infant mortality for Kenya from proportions dead among children born in the last 24 months before the 1999 census

	Births in the last 24 months	*Number of children dead*	*Proportion dead*	*Conversion factor*	*Estimated Infant Mortality (per thousand)*
Total	1 694 451	121 500	0.0717	1.09	78.2
Males	869 537	63 839	0.0734	1.09	80.0
Females	824 914	57 661	0.0699	1.09	76.2

Note: These estimates differ slightly from those published in the Kenya census report Vol. III. *Analytical Report on Population Dynamics*, Table 3.5, (Central Bureau of Statistics 2002) which used a conversion factor of 1.11.

Our estimate appears to be somewhat higher than those from the DHSs, but lies well within the confidence limits of the latter. If anything it lends weight to the belief that, in the era of AIDS, direct mortality estimates from birth histories are more likely to be too low than too high.

DISCUSSION

The procedure we have described has substantial advantages: it provides estimates which are more up-to-date than those from other sources; it is robust to the selectivity bias arising from the mother-to-child transmission of HIV; the required tabulations and the adjustments needed to convert the proportions dead into estimates of infant mortality are simple to make. But it is vulnerable to reporting errors. Census and survey reports on births occurring in the preceding year are frequently under-reported, and have to be corrected with procedures such as the P/F ratio or variations on that theme. Thus, as with all indirect methods, the results need to be examined critically in comparison with those from other sources. Sometimes the method works and sometimes it does not. As we have seen above, it gave good results with the 1999 Kenya census; but with the previous census of 1989 it gave an infant mortality rate of about 45 per thousand which was unacceptably low.

But if the necessary questions have been asked, there is no reason for not making the requisite tables and calculations. Demographers analysing data from developing countries need to explore every avenue and use every weapon at their disposal. The simple technique we have developed here makes a modest addition to their armament.

References

Artzrouni Marc and Zaba Basia. 2003. "HIV-induced bias in the estimation of child mortality using birth history reports". Paper presented at the Technical Meeting on HIV and Child Mortality. London, March 2003.

Blacker J.G.C. 1984. "Experiences in the use of special mortality questions in multi-purpose surveys: the single-round approach", in *Data Bases for Mortality Measurement*. New York: United Nations, 1984.

Blacker J. nd. *Indonesia Survey of Infant and Child Mortality: Report on the Principal Findings.* London: Centre for Population Studies, London School of Hygiene and Tropical Medicine.

Brass William, Coale Ansley J., Demeny Paul, Heisel Don F. *et al.* (eds). 1968. *The Demography of Tropical Africa*. Princeton NJ: Princeton University Press.

Cantrelle P. and Leridon H. 1971. "Breastfeeding, mortality in childhood, and fertility in a rural zone of Senegal". *Population Studies* **25**(3):503–34.

Central Bureau of Statistics (CBS) [Kenya], Ministry of Health (MOH) [Kenya], and ORC Macro. 2004. *Kenya Demographic and Health Survey 2003.* Calverton, Maryland: CBS, MOH and ORC Macro.

Central Bureau of Statistics, Ministry of Finance and Planning. 2002. *Kenya 1999 Population and Housing Census Volume III: Analytical Report on Population Dynamics,* Nairobi.

Chackiel Juan and Gough Hew. 1989. "Errors and biases in procedures for estimation of infant mortality from survival of the last-born child". *International Population Conference New Delhi* 1989, Vol. 2 pp.113–27.

Ewbank D.C. 1982. "The sources of error in Brass's method for estimating child survival: the case of Bangladesh". *Population Studies* **36**(2):459–74.

International Centre for Diarrhoeal Disease Research, Bangladesh. 1995. *Demographic Surveillance System – Matlab.* Volume 23: Registration of Demographic Events – 1992, Dhaka 1995.

INDEPTH. 2002. *Population and Health in Developing Countries*, Vol. 1 *Population, Health, and Survival at INDEPTH Sites.* Ottawa: International Development Research Centre.

Keyfitz Nathan. 1966. "Finite Approximations in Demography". *Population Studies* **19**(3):281–95.

Mahy Mary and Zaba Basia. 2005. "Measuring child mortality in AIDS-affected countries". In press, *Population Bulletin of the United Nations.* http://www.un.org/esa/population/publications/adultmort/Adultmortality.htm

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

APPENDIX

Sampling Errors of the Indonesian Mortality Estimates

Sampling errors were calculated on the proportions dead among last and next-to-last births, and on the proportions dead among children ever born by age group of mother, using data on the numbers of last and next-to-last births with numbers dead, and of children ever born by age of mother with numbers dead. The sampling errors (SE) are calculated according to the formula

$$SE = \sqrt{\frac{C(\sum y_i^2 - 2.P.\sum x_i y_i - P^2 \sum x_i^2)}{(C-1)(\sum x_i)^2}}$$

where y_i and x_i are the numerator (number of children dead) and denominator (number of births) in cluster i respectively;

P is the overall proportion dead $(\frac{\sum y_i}{\sum x_i})$;

C is the number of clusters.

The upper and lower limits of the estimated proportions (P) were then calculated as $P \pm 1.96$.SE. The random errors were calculated as

$$\sqrt{\frac{P(1-P)}{N}},$$

where N is the number of births. The design effect is the ratio of the calculated standard error (SE) to the random error, and thus measures the effect of using a cluster sample. The results are shown in Tables A1 and A2.

The effect of these margins of error on the estimates of infant and under-five mortality derived from the proportions dead among last and next-to-last births are shown in Table A3. The relative width of the confidence intervals for Yogyakarta, where the numbers of births were particularly small, is such that the discrepancy between the estimates of $q(5)$ derived from the previous births and those obtained from the other data sets can easily be explained in terms of sampling error.

Table A1 Sampling errors of proportion dead of last births

Province	*Proportion dead*	*Standard error*	*Upper limit*	*Lower limit*	*N*	*Random error*	*Design effect*
Yogyakarta	0.03712	0.00863	0.05405	0.02020	431	0.00911	0.948
Sumatera Selatan	0.05675	0.00736	0.07117	0.04232	793	0.00822	0.896
Nusa Tenggara Barat	0.09964	0.01160	0.12237	0.07691	833	0.01038	1.117

Table A2 Sampling errors of proportions of children dead by age group of mother

Province group and age	*Proportion dead*	*Standard error*	*Upper limit*	*Lower limit*	*N*	*Random error*	*Design effect*
Yogyakarta							
20–24	0.04815	0.01447	0.07650	0.01979	270	0.01303	1.110
25–29	0.03443	0.00670	0.04756	0.02131	697	0.00691	0.970
30–34	0.04826	0.00997	0.06780	0.02873	1409	0.00571	1.746
35–39	0.07562	0.00878	0.09283	0.05840	1706	0.00640	1.372
40–44	0.07433	0.00744	0.08890	0.05975	1601	0.00656	1.134
45–49	0.08971	0.00942	0.10817	0.07124	1516	0.00734	1.284
Sumatera Selatan							
15–19	0.14815	0.03952	0.22562	0.07068	81	0.03947	1.001
20–24	0.12825	0.01484	0.15734	0.09916	577	0.01392	1.066
25–29	0.11084	0.01038	0.13118	0.09050	1633	0.00777	1.336
30–34	0.10460	0.00838	0.12104	0.08817	2476	0.00615	1.363
35–39	0.12980	0.00960	0.14862	0.11098	2681	0.00649	1.479
40–44	0.14828	0.01160	0.17102	0.12554	2003	0.00794	1.461
45–49	0.18607	0.01641	0.21823	0.15391	1967	0.00877	1.870
Nusa Tenggara Barat							
15–19	0.11236	0.02866	0.16852	0.05620	89	0.03348	0.856
20–24	0.15919	0.01555	0.18966	0.12872	691	0.01395	1.117
25–29	0.18509	0.01330	0.21116	0.15902	1556	0.00985	1.351
30–34	0.21707	0.01438	0.24525	0.18888	2414	0.00839	1.714
35–39	0.24072	0.01217	0.26457	0.21687	2559	0.00845	1.440
40–44	0.27891	0.01541	0.30912	0.24869	2266	0.00942	1.636
45–49	0.29807	0.01959	0.33646	0.25968	1815	0.01074	1.824

Table A3 Upper and lower limits of estimates of infant and under-five mortality from proportions dead among last births

Province	*Infant mortality*		*Under-five mortality*	
	Upper limit	*Lower limit*	*Upper limit*	*Lower limit*
Yogyakarta	57.3	21.4	73.2	27.6
Sumatera Selatan	75.9	45.1	129.1	78.6
Nusa Tenggara Barat	133.0	83.6	216.9	141.4

SAJDem **10**(1&2):43–68

Population mobility and household dynamics in rural South Africa: implications for demographic and health research

Victoria Hosegood[1], Justus Benzler[2], Geoff C Solarsh[3]

Abstract

Rural areas of South Africa are characterised by highly mobile populations with high rates of circular migration and fluid household and social relationships. The health and demography of South Africa is undergoing substantial changes as a result of the rapidly progressing HIV epidemic. Consequently, for researchers and policy makers, high quality longitudinal socio-demographic and health data from affected areas is critically important. Sociologists and anthropologists have long commented on the conceptual limitations of the social data collected by demographic and health surveys, censuses and demographic surveillance system data.

This paper describes a study of households in rural KwaZulu-Natal, South Africa, that used qualitative approaches to inform the design of a large, demographic surveillance system. Using in-depth interviews with members of 60 households this research examined local descriptions and categorisations of social structures and processes such as households, residency, intra- and inter-household relationships, parenting, partnerships, and household authority. Thirty-five percent of the total household population were non-resident, 41 per cent of all adult members and 29 per cent of all child members were non-resident. We compare our findings with those obtained by applying standard definitions used in most other demographic and health surveys and censuses. We discuss the research potential in modifying population-based data collection instruments to model the complex and dynamic nature of social arrangements in rural areas of southern Africa.

Keywords

Demographic surveys, migration, households, South Africa

---

1 Corresponding author: Africa Centre for Health and Population Studies, Faculty of Health Sciences, University of KwaZulu-Natal, PO Box 198, Mtubatuba, 3995, South Africa. Centre for Population Studies, 49–51 Bedford Square, London, WC1B 3DP, UK. Email: *vicky.hosegood@lshtm.ac.uk*

2 Robert Koch-Institut, Abt. Infektionsepidemiologie, FG Surveillance, Berlin, Germany.

3 Monash University School of Rural Health, Bendigo, Victoria 3552, Australia.

INTRODUCTION

A survey team conducts a demographic and health household survey in a rural area of KwaZulu-Natal, South Africa. Having entered a homestead and introduced herself to the only adult present, a fieldworker starts by trying to identify who is the head of the household. Her informant says that her husband is the head of the household. He is working in Durban and returns for a few nights at the end of each month or when the need arises. Having been trained that a household member must be someone who spends most nights of the week at the homestead, she decides to omit him and record the head's wife as the head of the household instead. The fieldworker then asks about the other household members, and lists two of the couple's children but not their teenage son who stays with his aunt during the week in order to be close to school. The fieldworker also does not record information about a young man who has lived with the family for many years in order to look after their cattle because the wife had not mentioned him when asked to list the household members.

The data collected by the fieldworker thus characterises this household as being headed by a woman and having three members (a woman and two children). The physical presence and the roles played by the respondent's husband, her schoolboy son and the herdboy would not be captured, neither would the events that occurred to them. In addition, by asking only about the household of the first informant she meets, she has completely overlooked another young couple living at the same homestead, who rent a shack and consider themselves to be a separate household.

The challenges faced by this fieldworker in translating complex social relationships and residential arrangements into the definitions and data items used in the survey will be recognisable for anyone who has conducted population-based studies in southern Africa. All censuses, demographic and health surveys (DHS) and demographic surveillance systems (DSS) collect information about households, for example, household composition, head and assets, and these attributes are often examined as potential correlates of health and welfare outcomes. Most use a *co-residential concept of households*, also referred to as a *household dwelling concept* (de Jong Gierveld 1995). South African surveys and censuses primarily use a co-residential, *one pot* concept of a domestic unit.

The 1996 October Household Survey, a national economic, health and demographic survey of South Africa instructed its fieldworkers that:

A household consists of a person or group of persons who:
I. eat together and share resources; and
II. normally resides at least four nights a week at the specific visiting point;

For the purposes of this survey, a live-in domestic worker is considered to belong to a separate household.

The 1998 South Africa Demographic and Health Survey defined a household as:

... a person or group of people living together, eating together and sharing resources.

The definitions of household used in most population-based studies differ markedly from the characterisation of households by anthropologists and sociologists (Fortes 1966; Spiegel 1986; Burch 1995). In contrast, anthropologists and sociologists describe southern African households as dynamic and complex. High levels of adult economic migration involving migration from rural areas to employment opportunities both in urban areas and rural areas (mines, farms, game reserves) is well-established (Berry 1993). The apartheid policies entrenched a widespread pattern of *circular* migration (Spiegel 1980; Murray 1981; Leliveld 1997). These policies included the prohibition of free movement of black South Africans, particularly of black women; the creation of designated homelands (*bantustans)* for these populations in rural areas; and restrictions on family members accompanying mainly male migrant workers (Schapera 1947; Murray 1980). Despite the end of apartheid legislation, circular migration has been sustained by differentials in economic opportunities, customary and state land tenure systems, the cost of living and living conditions in urban areas (Lurie, Harrison, Wilkinson *et al*. 1997). Circular migrants move between their rural households and their place(s) of employment, typically making return visits to the rural areas during holidays, emergencies or when employment ends; supporting rural households through remittances of cash or goods. Women form an increasing proportion of the migrant population (Hill, Hosegood and Gafos 2005).

Anthropologists and sociologists also characterise southern African households as highly fluid and adaptable domestic arrangements. There are frequent exchanges of individuals between households, often in response to labour migration, financial insecurity, ill-health, and death (Niehaus 1994; Ross 1996; Spiegel, Watson and Wilkinson 1996; Townsend 1997a; Ross 2003; Russell 2003b). Child fostering patterns in west and southern Africa are, at least in part, adaptive strategies to optimise individual and household survival (Goody 1982; Bledsoe 1995; Van der Waal 1996).

Papers by Margo Russell provide a valuable summary of the extensive literature on households and social relationships in southern Africa, and discuss the way in which the widely used definitions and classifications of black households have been influenced by the application of models of the household and family from western patterns (Russell 2003b; Russell 2003a).

In the light of the extensive body of literature about historical and contemporary social organisation in southern Africa, the validity of survey data on social structures has been questioned. These include discussion about definitions of households (Keilman 1995; Spiegel 1996), household headship (O'Laughlin 1998; Budlender 2003), conjugal relationships and marriage (Bledsoe and Pison 1994; Locoh 1994; Jones 1998; Budlender, Chobokoane and Simelane 2005), and parenting (Jones 1993; Bledsoe 1995; Townsend 1997a). A broader critique, levelled at demographers in particular, has been that there has been insufficient critical reflection about the collection, analysis, and interpretation of socio-demographic data (Caldwell and Hill 1988; Greenhalgh 1996; Kertzer and Fricke 1997).

Given both the evidence and the criticisms, why have censuses and demographic and health surveys continued to use a co-residential definition of household in southern Africa? There are several reasons for its use. Living together is considered a marker of domestic ties in the social group, facilitating domestic activities such as eating, sharing of resources and a common environment. There are also methodological and practical reasons. Study or census populations are usually linked through residence to a defined area (country, village, study area). Individuals who live outside this area are not eligible for inclusion. In censuses and DSSs, reliable estimates of population denominators and demographic rates require low rates of double-counting of individuals or missed events. Thus, in many cross-sectional survey methods, the approach is to only record people present on the night of the census or interview (the *de facto* approach).

In longitudinal population-based surveys such as DSSs, individuals and households are observed over time. Unlike cohort or panel studies that can opt to follow subjects wherever they live, most DSSs observe individuals only while they are resident in the surveillance area. Most DSSs will track an individual who migrates within the surveillance area, however observations stop if he/she migrates out the study area. Longitudinal population-based surveys opt for a *de facto* population on the grounds that only the resident population is ‘at risk’ of having demographic events in the surveillance area, for example births, deaths, or using a health clinic or school (Garenne 1997; Indepth Network 2002). Lastly, because the co-residential definition of a household is widely used, new studies use the same approach in order for their data to be comparable with other surveys.

In 1997, the Africa Centre for Health and Population Studies (ACHPS) began developing a DSS known as the Africa Centre Demographic Information System (ACDIS). Starting in January 2000 it has prospectively followed up a population of approximately 85 000 people living in 11 000 households in rural KwaZulu-Natal, South Africa (Solarsh, Benzler, Hosegood *et al*. 2002; Hosegood and Timæus 2005). This paper describes exploratory research conducted in sixty households during the development phase. One of the most important objectives for starting ACDIS was to measure the social, demographic and health consequences of a rapid and severe HIV epidemic, and to provide a population platform for intervention research. Therefore, ACDIS would need to use a conceptual model and data collection instruments that could capture and represent complex information about people, their social groups, their places of residences, and the events that happen to them; in a way that closely reflects the social, economic and infrastructural realities for this population.

As part of developing the conceptual model and instruments we conducted an exploratory pilot study and collected information about the perceptions of household respondents and fieldworkers about a range of socio-demographic subjects and processes including household, membership, place of residence, migration, partnership and co-habitation. During repeated visits to study households, we tested different ways of collecting socio-demographic data including diagrams, semi-structured question guides, and field notes. In this paper we discuss how these concepts were perceived by respondents and the degree of consistency in reports by different respondents within the same household. We also discuss how these emic characterisations map onto the survey definitions commonly used in South Africa. The findings from this pilot work were subsequently incorporated into the ACDIS conceptual model, definitions, data collection sequencing, and structured questionnaires, and these have been described elsewhere (Hosegood and Timæus 2005).

STUDY LOCATION

This study was conducted in the former Hlabisa district in KwaZulu-Natal, South Africa, 250 km north of the regional capital of Durban. The district had a resident population of approximately 200 000 (Curtis, Bradshaw and Nojilane 2002), almost exclusively Zulu-speaking. Most people live on Zulu tribal authority land, with around 10 per cent living in two townships administered by a local municipal council. In the tribal area there are no identifiable villages, and homesteads are scattered, separated by grazing, arable or unused land. Most homesteads in the tribal area are without electricity, piped water or sanitation, in contrast to the townships' well-developed infrastructure. Despite being a rural area, the principal source of income for most households is waged employment and state pensions rather than agriculture (Marcus 1998; Case and Ardington 2004).

METHODS

The pilot study enrolled 60 households who were visited between January and March 1998. To ensure geographical and economic representativeness of study households, spatial co-ordinates were calculated for the middle of each of the 12 clinic catchments within the district. At this place, fieldworkers visited the closest homesteads to conduct a neighbourhood mapping exercise. Accompanied by local key informants, they walked through the informant-defined local neighbourhood recording and mapping information about each homestead and the resident households. Informants were also asked to rank households by economic status (wealthy, middle or low) relative to other households in the neighbourhood. A sample of five households proportionate to the wealth ranking distribution was selected in each of the 12 neighbourhoods (n=60) (Hosegood 1998; Marcus 1998).

Repeated visits were made to each household to collect different types of data (household definition and composition, residential mobility, and social, conjugal and parental relationships). Different household respondents were interviewed either collectively or individually to examine consistency of reporting, reliability and quality of contemporary and historical information about the household and its members. We also used a variety of different methods of data collection, from observational notes recorded by field-workers, genograms, structured questionnaires, and indepth interviews with individual household respondents. This provided opportunities to triangulate information from respondents within the same household and from the same respondent to questions asked in different ways. We tried to involve as many household members in the exercises as possible. Respondents were noted in each data collection exercise for comparison. In several households, contact with adult members was difficult during the day. In these households information was collected from older children but compared with information from adults collected at weekends.

The genogram was used as the first step in collecting information about household composition. This tool allows a visual representation of social relationships and information to be created with one or more respondents and discussed in an interactive manner (Cross and Preston-Whyte 1985; Watts and Shrader 1998). Simple coding conventions were developed to ensure comparability between genograms, such as double line between two people meaning that they are married. Initially household respondents were asked to describe which people they considered to be members of their household, and to provide information about their age, sex and relationship to each other. This included current and past relationships, so many genograms included deceased relatives or ex-partners, included by respondents to explain why certain people were members of the household. Other people resident at, or visiting the homestead, were then added to the genogram with indications about their social inclusion or exclusion. In a semi-structured questionnaire, information about all people appearing on the genogram was collected about a range of characteristics: marital histories, conjugal relationships, residential patterns, frequency and duration of stays with the household, their children (survival status, current residence, contact), the arrangements for caring for children in the household, employment and education histories, and connections to other households in the neighbourhood. After the initial interview notes had been collated, the genograms were discussed again with household respondents. They were asked to re-identify the members of their household and to explain their reasons for including or excluding particular people. Additional notes were recorded by fieldworkers to supplement the information in the diagram, as well as observational comments about discussions that occurred during the process. Figures 1 and 2 show two examples of completed genograms from the study (all names have been changed).

In a second phase, a semi-structured questionnaire was administered to household respondents to collect information about household members, their relationships, and residential patterns. The information was collated and a summary description was generated for each household and compared with the information from the genogram. The characteristics of the 60 households are described in this paper.

RESULTS

General characteristics of household members

Table 1 presents characteristics of the 60 households, a total study population of 752 individual household members. The average household size was 12.5 members but varied considerably (SD=7.2, min 2, max 38). One third (34 per cent) of all household members were non-resident. An equal sex ratio was found for children with slightly more adult women (52 per cent) than men. The sex ratio and the age distribution of the self-defined household population were therefore not distorted by the high rate of adult male labour migration.

Table 1 Summary statistics for all members of the 60 study households

Characteristics of household members	*Residents*		*Non-residents*		*Total per cent*	*All members*	
	Number	*Per cent*	*Number*	*Per cent*		*Number*	*Per cent*
Adults (18+ years)							
Male	79	45	96	55	100	175	48
Female	135	72	53	28	100	188	52
Total	214	59	149	41	100	363	100
Children (<18 years)							
Male	140	71	58	29	100	198	51
Female	137	72	54	28	100	191	49
Total	277	71	112	29	100	389	100
Total members	491	65	261	35	100	752	

Distinction between homestead and household

In mapping the tribal area neighbourhoods, informants made a clear distinction between the physical dwelling 'homestead' (*umuzi*) and its associated social group(s) 'households' (*umndeni*). Homesteads are demarcated plots allocated to a single individual by an *induna*, the chief's local representative. Homesteads can encompass domestic buildings as well as shops, businesses and agricultural land and several households may occupy a single homestead. Four common types of household arrangements at homesteads were observed.

(i) Single households occupying one homestead.

(ii) Two or more households sharing a homestead acknowledging different household heads. These heads were often kin, typically brothers or father and his sons.

(iii) Several separate households associated with a polygamously married man. Although living at the same homestead and acknowledging him as the head, each wife considered herself and her children to form a separate household.

(iv) Several single or groups of tenants paying rent to the homestead owner. The owner's household was often also living at the homestead.

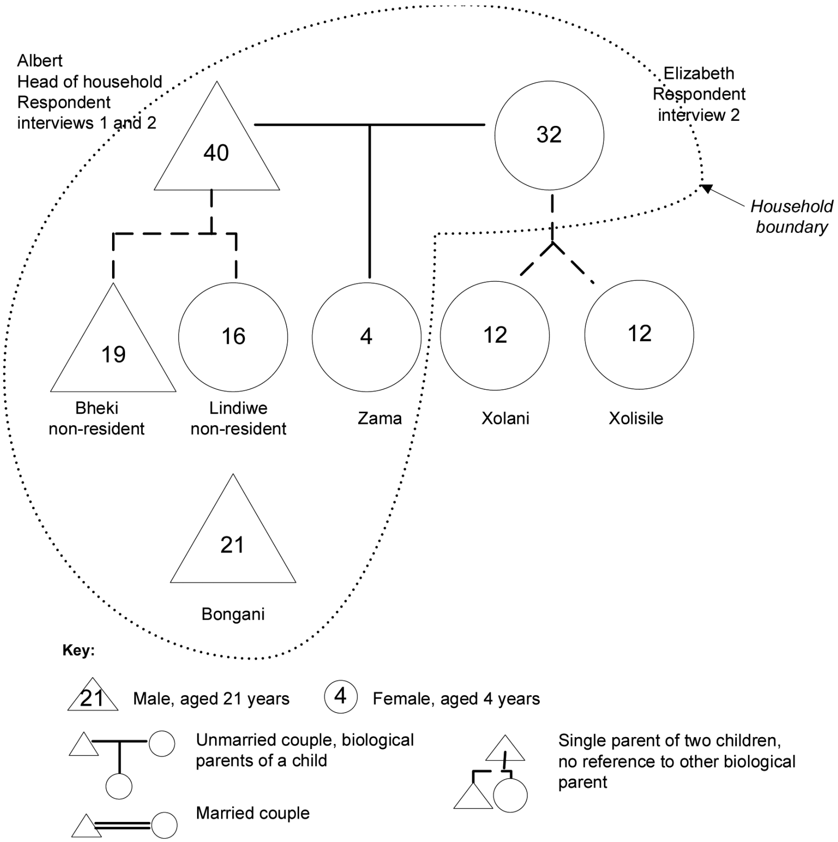
Distinguishing between residential dwelling (homestead) and the resident social group (household) influenced the household selection process. Whilst homesteads could be identified from maps or be directly observed, identifying the households resident at the homestead could only be done through interviews.

Defining household membership

Consensus between household respondents about who were members of their household was readily achieved. Households were described in terms of kinship, the recognition of a single head of household, and shared responsibilities and assets. The way in which most respondents constructed the genogram was suggestive of their perception of the household. Most began by listing the head of household and his/her closest relatives such as spouse, children, parents, before describing other kin such as adult siblings and grandchildren of the head. Distant kin or unrelated members were added last. Household membership was not seen as contingent upon the person's place of residence.

The majority of households could be classified as both laterally and three-generationally extended households, i.e. households that include the children and grandchildren of the head of household, as well as other relatives of the head and non-related individuals (Gage, Sommervelt and Piani 1996). Few households had an elementary structure, defined as households consisting only of parents and their biological children. Where households were elementary they were mostly single-parent households rather than nuclear households or polygamous households.

Figure 1 Genogram of the Dube household



When viewed statically at one point in time, these household typologies, based largely on kinship structures, hide changes in social connectedness that have occurred in the past. For example, one fostered child without kinship relations with the other household members had joined the household after her mother, a former colleague of the head of household's wife, had died. Household composition was frequently explained in terms of people who were dead or were no longer present. This was common where women or men were living with non-biological children who were step-children of partners who were not included as household members because they had died, had been abandoned by, or left their partners.

Figure 1 Notes to genogram of the Dube household

The genogram alongside was created by Albert at interview 1 and discussed with Albert and Elizabeth at interview 2.

Albert and Elizabeth have been a child Zama who is 4 years old but have only been living together for two years. They are not married but some lobola has been paid. Albert is unemployed, Elizabeth occasionally earns money selling vegetables. The Dube household has six members.

Four of whom are resident most nights: Albert, Elizabeth, Zama and Albert's nephew, Bongani. Bheki and Lindiwe, Albert's children from a previous relationship, are also considered by both Albert and Elizabeth to be members of the household. However, they live most of the time with their mother in a neighbouring town (one and a half hours away) but stay with their father during school holidays. They live with their mother in term time because she lives closer to their school. Albert considered them to be **part of his household** because they are his children, he supports them, and because they regularly come for extended visits. Elizabeth also has two children by a previous relationship, twins Xolani and Xolisile. However, neither child is considered by Albert or Elizabeth to be a member of this household. From an early age they have lived with Elizabeth's parents in another province. Elizabeth visits them twice a year but they have never visited her while she has been living with Albert. Elizabeth said that they are members of her parental household but not a part of **her and Albert's** household.

Bongani is Albert's nephew. He joined the household as a young child when his father died.

Respondents described a myriad of circumstances in which people had joined the household. In Figure 1, the household described by the genogram included a 21-year-old son of the household head's deceased brother who had joined the household many years previously upon his father's death. In several households children had moved in with a notional role as domestic workers or cattle herders (Bray 2002, 2003). Although initially they were not considered members, the children had eventually been assimilated into the household over time and supported in schooling or marriage as recognised members. A high percentage of households included children whose mother and/or father was not also a household member. Only 30 per cent of children were members of the same household as both their parents. The social, economic and practical reasons determining the social and residential mobility of children were extremely diverse and, in explaining the circumstances respondents needed to give detailed information about events occurring in, or relationships between, several other households.

In a small number of cases residents were not considered to be members of any social group at the homestead, rather they were seen as being just single people or belonging to households somewhere else. These were often single adults or children who stayed with the main household, often sharing the same buildings, food and assets. The relationship of these people to the resident households included domestic or farm workers working for the household either for payment or food and shelter; kin or friends staying because of work, schooling or problems in their own families. These individuals were considered to be members of other households elsewhere and many returned to other households for visits or long-term.

Residency status of household members

Respondents described a person's place of residence as the place where they are based rather than the place where a person is currently staying. Comparing *de facto* data collected at the first visit with self-defined residency status suggested that respondents were distinguishing non-resident members who were visiting during their holidays from resident members who were currently away. These results present the residential status of individuals at the homestead as reported by respondents. More adults were non-resident (41 per cent) than children (29 per cent), and a higher proportion of men (55 per cent) were non-resident than women (28 per cent) (Table 1). For children no sex differences in residential status were observed. Non-residency was a ubiquitous feature of households, with 88 per cent of households reporting at least one non-resident adult and over half of all households (53 per cent) had at least one non-resident child member (Table 2).

Table 2 Household level characteristics

Household composition	*No of households*	*Per cent*
Households with a resident adult	59	98
Households with a non-resident adult	50	88
Households with a non-resident child	32	53
Households with a male head	43	72
Households with a resident head	41	68

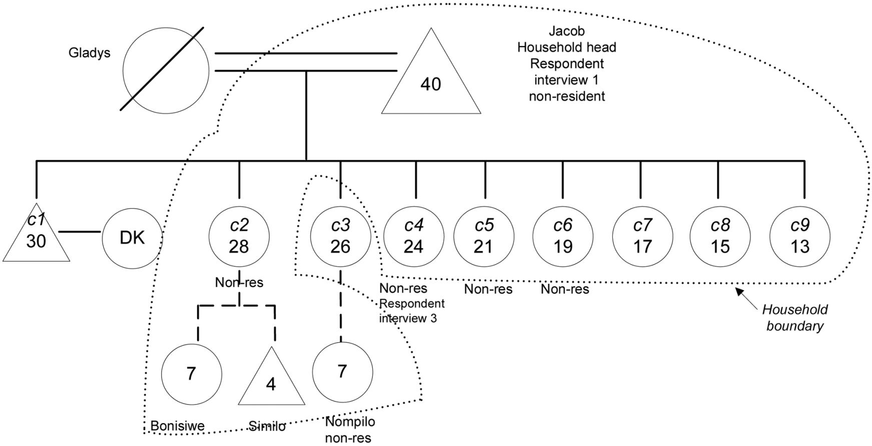
The main reason for adult migration was either having or seeking employment, predominantly within KwaZulu-Natal. Although most worked in the larger coastal towns and cities, there was a wide range of other work in rural areas such as commercial agriculture, forestry and game tourism. Long distance migration was more common for men than women with 18 per cent of non-resident men resident in Johannesburg, Gauteng Province, compared with only 4 per cent of women. This is also reflected in the main types of Gauteng employment reported: mining, production and security. Women and children were more likely to migrate locally within the district. The household described in Figure 2 illustrates the diversity of types of labour migration and the range of return patterns. Labour migration often indirectly promoted the mobility of other members. For example, non-resident women and children accompanying their migrant husband/partner or father, or adolescents and adults moving, by command or choice, to live with a labour migrant who could provide better financial support, security or access to employment and education opportunities. Of importance for understanding migration patterns in South Africa was the observation that many migrants, although working in urban areas, had accommodation in peri-urban settlements or surrounding rural areas.

The main reasons reported for children migrating from their parental household(s) were schooling and/or, to be cared for by another household. A prevalent pattern of child migration was that of oscillating migration between two homesteads with the concomitant result that the child was recognised as a member by two households. The need to live in one place during school terms and another during holidays was a common reason for this. However, more complex patterns of child mobility and social arrangements were generated by labour migration inhibiting the co-residence of the child's parents, high rates of partnership and marital instability, and cultural norms that discourage co-habitation of unmarried partners.

Pattern of visits by non-resident members to households

In general non-resident members were expected to return whenever possible and for family events such as funerals and marriages; and most did so. Migrants were also described as making return visits due to pregnancy and illness since many labour migrants have restricted living arrangements and limited social networks in the towns and cities. Female labour migrants typically returned to rural households late in pregnancy for support during and after birth. As many were unable, or choose not to, take their children with them on returning to work, these children were in the care of other household

Figure 2 Genogram of the Dlamini household



This genogram was created by Jacob and c2 at interview 1. The genogram was later discussed with c4.

Jacob is the resident head of this household. His wife, Gladys, had died six months earlier. Together they had nine children. The household now consists of eleven members, Albert, seven of his daughters, and three grandchildren. Several of these members are living most of the time somewhere else but return regularly.

c1, the oldest son, has established a separate household with his girlfriend in Durban. They have a house in Durban but he actually lives and works in another city and he joins his girlfriend once a month. He is not considered a member.

c2 works in a neighbouring town and comes back once a week for a night. Her two young children stay with the household. The youngest, is cared for during the day by her sister c3 who is not a member but lives nearby.

c3, a 23-year-old, was not considered to be a household member. She lives with her boyfriend in a local township. All informants questioned agreed about c3's non-inclusion in the household membership list; a general disapproval for her lifestyle was given as the reason – not looking after her child herself and going to live with a new boyfriend.

In contrast, c3's 7-year old child, Nompilo, is considered a household member even though she lives most of the time with her father (not c3's current boyfriend). This place is close by and there is a lot of contact between Nompilo, her aunts and her cousins.

c4 teaches in a town about four hours away and comes back once a month.

c5 is a university student in Durban and comes back for holidays and some weekends.

c6 is a shop assistant in a town about three hours away and comes home whenever she has a day off.

c7,8 and 9 live with the household and go to school locally.

members. Several members described as non-resident members had been staying at the homestead for an extended period prior to the fieldworker visit because of ill-health. In such situations, the person themselves or other respondents, often found it awkward to describe their stay as anything more than a temporary visit until their health improved. Where there was no expectation of recovery, the duration of illness had been long, or the person had lost their job because of ill-health, former migrants being more likely to be described as being resident.

To test the methodological assumption that non-residents are not exposed to risk factors in the study area, information was collected about the frequency and duration of their return visits. Table 3 presents the visit patterns for non-resident adult members. 50 per cent returned at least once a month for one or two days and an additional 30 per cent spent longer holidays such as Christmas and personal leave with their rural households. For some (7 per cent) it was not yet clear how often they would be returning since they had only recently left often to look for work. The pattern of return visits by non-resident children and women was more difficult to categorise. With the exception of children and women who were non-resident because of formal employment or schooling, the types of reasons given for their migration such as temporary work, caring for other relatives, accompanying someone else, created less predictable patterns of presence and absence.

Table 3 Return patterns by non-resident adult members (n=149)

Return patterns	*No*	*Per cent*
Weekends	24	16
Monthly	51	34
Holidays	45	30
Infrequently	10	7
Have not returned for >2 years	9	6
Recently left no pattern established	10	7

Membership of multiple households

Some members of study households were described as belonging to more than one household including polygamous men whose wives maintained separate households, children living away from their parental households, and domestic workers whose familial households were elsewhere. In five cases fieldworkers visited the other non-study household to establish their perceptions of the individual's membership status. One was a boy of 12 years who was reported to be a member of a study household as well as another in the neighbourhood. The boy had lived in a household of non-relatives for several years; however, he also made frequent visits to his parents' household although he did not sleep there. In this case, as in the others examined, respondents from the different households considered the child to be a member of more than one household.

Household authority

Like other sub-Saharan societies, rural South African households recognise a single individual, usually an adult male, as having a senior position of authority in the household. The majority of households had an adult male head of household (72 per cent), and 32 per cent of all household heads were non-resident. In two households the oldest woman reported her husband as the head of household even though they were estranged. No child-headed households were present in the sample. The roles of the household head varied widely in the study households depending on their marital status, sex, age, residential status and household composition.

An additional step of data processing was to classify the relationship of household members to the head of household using standard DHS codes (Demographic and Health Surveys 1990). In many of the households we experienced difficulty in consistently applying the codes 'other relative' and 'not related'. Relationships were often described in relation to the household head's spouse or siblings, for example, his wife's brother's son, or, his brother's wife's daughter. The DHS codes were inadequate to capture many of the complex and overlapping relationships documented in the genograms.

DISCUSSION

In the neighbourhood mapping exercise and household selection, a clear distinction was made between the homestead as a fixed geographic location and a household as a social group whose members may have residence in different places at any time. A single homestead may be the place of residence for several households. Studies often do not make this distinction explicit, conflating homestead with household in their approach to identifying households in the field. This risks introducing a selection bias with smaller, less socially prominent, or transient households being under-reported (Leliveld 1997). In the South African census, a distinction is made between a 'dwelling unit' and a household. Enumerators can record more than one household as living at the same unit.

The reasons respondents give as to why they considered some people to be members of their household and others not were very different from the co-residential and *one-pot* eating arrangement used in the South African DHS (South Africa DHS 1999). Rather than co-residence, householders emphasised their sense of belonging together and maintaining social bonds through responsibility for, or dependence on, other household members. Members of the same household shared an understanding of the authority hierarchies within the household. Most members were related to each other through kinship though many households reported child and adult members who were unrelated to other people in the household. Non-residents constituted a significant proportion (35 per cent) of all household members. The phenomenon of non-resident household members was prevalent across all socio-economic levels and geographical areas.

The finding that community household membership is independent of a person's place of residence is one supported by anthropological and sociological studies in South Africa and other countries in the region. Spiegel, Watson and Wilkinson (1996) found that household respondents in the Cape Town area of South Africa did not use a definition of 'eating from the same pot' when describing the members of their household. Ethnographic research in Lesotho and Botswana, based on self-reported household composition also identified migrants who were not co-resident with the majority of other members (Murray 1981; Townsend 1997b).

The high rate of population mobility shown in this study is also reported by other studies in rural areas of southern Africa (Murray 1978; Murray 1980; Spiegel 1986; Timæus and Graham 1989; Ross 1996). Although the focus of previous studies has primarily been on adult migration, similarly high levels of child mobility have also been reported. In a study of children in a settlement in the Northern Province (now Limpopo), South Africa, 41 per cent of children changed their place of residence at least once in a single year (Van der Waal 1996). The movement of people, particularly children, between households has been described as an important coping strategy in response to financial constraints and parental absence or orphanhood (Izzard 1985; Ross 1996).

Migrant members maintained their social ties to the household by periodic return visits, financial and material contributions to, or dependency on, other household members. While those who frequently returned or made contributions consolidated their membership through their involvement with the household, many households included non-resident members who had not made a visit for more than a year or two. Respondents gave a variety of reasons for their inclusion, for example, where they felt that circumstances had prevented them from visiting, or because the person had a special role in the household, for example, he or she was the head of the household or was the oldest son. The membership status of these long-term absentees generated the greatest amount of disagreement between respondents from the same household, particularly from those members who felt the non-resident had abandoned the household. We were unable to triangulate these discussions with the perception of non-residents themselves. In Ferguson's study of migrant labourers in Zambia, men moving between rural and urban areas develop complex social identities influenced by factors including the degree of permanency at their destination (Ferguson 1999).

Thus, the pattern of social organisation that exists in this area is at odds with the standard survey definition of a household as a co-residential domestic group. Applying the 1996 OHS and DHS definition of a household as cited in the introduction would have generated a very different household population. 41 per cent of adult and 29 per cent of child household members would have been excluded from the study population. Among those who would not have been eligible on the basis of non-residency would have been 32 per cent of all household heads.

What are the implications for demographic and health research of excluding non-residents from the study population? Social processes such as household composition, migration, conjugal and parental relationships play an important role in determining demographic, health and economic outcomes. Excluding migrants would alter the demographic profile substantially and limit the collection of data about important relationships between individuals. Many linkages could no longer have been made for many resident members such as parent–child pairs, spousal partner pairs, and the relationship of members to their self-defined household head.

The implications of representing these self-defined social units and relationships can be illustrated by considering three research topics: children's living arrangements, migration and HIV/AIDS, and poverty and health. Given the high levels of adult mortality due to AIDS, there is considerable interest in determining the impact of parental illness and orphanhood on the living arrangements of children. In our study, one third of all households included a child for whom neither parent was a member. These results are similar to those reported by Gay (1980) in south-western Lesotho where 34 per cent of households in one village included one or more resident children whose mothers or fathers were not also members of that domestic unit. In our study, non-marital fertility, adult migration, divorce and child fostering were identified as circumstances that generated frequent, oscillating migration of children between households both in the local area and beyond. In addition, our study found several children with concurrent membership of more than one household, often related to separate living arrangements of their parents and/or schooling. Such children have a complex set of multiple household environments that may moderate or exacerbate the impact of parental illness or death.

Applying a co-residential household definition would have excluded many non-resident parents of children. The father–child links would not have been for around 70 per cent of all children because although their fathers were members of the same household they were resident elsewhere. We can anticipate different levels of involvement with their children (financial, emotional, physical) for mothers and fathers who are socially linked to their children through a common household, whether resident or non-resident, and parents who share no social connections with their child.

The relationships between non-resident and resident members of the same household are also potentially important determinants of health outcomes. Circular migration has been identified as an important determinant in the rapidly progressing HIV epidemic in rural areas of Africa (Nunn, Wagner, Kamali *et al.* 1995; Lurie, Harrison, Wilkinson *et al.* 1997). The mobility of couples also influences fertility, use of health services and childcare arrangements. In addition to an improved understanding of transmission dynamics, information about the characteristics of sexual relationships of migrants and their rural partners can inform the design of health promotion strategies. Examining these processes in populations with high residential mobility requires data about non-residents that are unavailable from most rural household surveys (Booysen 2003).

Studies seeking to examine the role of rural household livelihoods and socio-economic status in determining health are also strengthened by accurate data on households, intra- and inter-household relationships and resource flows between households. Remittances of money and goods from migrants are a major determinant of household livelihoods in rural areas where local economic opportunities are limited. For migrants this connection to a rural household might reduce his/her own personal finances but may be exchanged for other kinds of support such as rural household members caring for the migrant's child or looking after the buildings, land and livestock in which they have invested. Interpreting household level economic data without information about migrants' contribution to, and share of, household assets may lead to poor or erroneous interpretations of household level data.

Our qualitative study also challenged the widely held assumption that only *de facto* household definitions should be the basis of eligibility in surveillance populations. The justification that non-resident individuals are not at risk of demographic events within the area is countered by our findings. The frequency of return visits was high for many non-residents, 50 per cent of adults returned for several days each month, and many also spent school holidays and annual leave with the rural household. Such return patterns result in considerable exposure time each year for many non-residents within the surveillance area. In addition, pregnancy, birth or illness often prompt return visits by non-residents because rural households are a major source of support. A resident-only population denominator would therefore not be consistent with the births and deaths that occur in the study area and may bias fertility and mortality estimates.

This study also suggests that many studies may be poorly recording resident people who live intimately with a household but are not considered by respondents to be a part of the core domestic unit. We identified domestic or farm workers, tenants, child and adult boarders and visitors living with households that did not spontaneously report them when describing the household. There is a need to identify such resident individuals to avoid under-enumeration, though the approach to recording their household affiliation is often unclear. Sometimes they are registered as separate households, in others they are assigned to the household with whom they live. Both approaches are problematic. Defining such a person as belonging to a separate household will elevate the number of single person households, as well as child- or women-headed households. Arbitrarily assigning them as members of the main household will mask their socially marginal status.

The finding that some people belong to more than one household concurrently has been given little attention in the southern African demographic literature. An ethnographic study of 30 residential sites in the Transkei, South Africa described similar patterns of multiple membership (Spiegel 1986). Given highly fluid households, high rates of circular migration, and the common practice of child fostering, it is not however an unexpected phenomenon. Adult migrants may belong to households in a rural area as well as at their migration destination. Migrants are sometimes accompanied by their spouse and/or children, or are joined by other members in order to seek work or financial support. Spiegel (1986) uses the term 'stretched household' to describe households where members reside in different places. The standard approach to assigning individuals to households in surveys or surveillance systems is to record only one household membership at any point in time, usually where they are reported as resident (Das Gupta, Aaby, Garenne *et al.* 1997; Garenne 1997; Binka, Ngom, Phillips *et al.* 1999). This approach limits the ability to present a comprehensive picture of social dynamics and the multiple domestic environments with which individuals interact.

This study has shown some of the complexity of social relationships and residential mobility in a rural area of South Africa. Population-based studies, whether cross-sectional surveys or longitudinal surveillance systems in southern Africa face the challenge of collecting data that represent more closely social and demographic contemporary realities in order to answer important questions about demography and health, particularly in the era of HIV/AIDS (Booysen and Arntz 2003). We suggest that longitudinal population-based studies, such as demographic surveillance systems can accommodate a more *emic* description of social structures without compromising data quality and comparability. The main modification required is to adapt household eligibility criteria to recognise the number and importance of non-resident adults and children in rural households.

Acknowledgements

This research was supported by grants from the Wellcome Trust, UK to the Africa Centre for Health and Population Studies (#50534), ACDIS (#50534). V. Hosegood was supported through a Wellcome Trust postdoctoral fellowship (#61145).

We would like to thank those involved in the original conception of this study: E. Preston-Whyte, D. Scott, T. Marcus, A. Harrison, and the late D. Le Sueur. The authors specifically acknowledge the contribution of M. Lurie to the study design and fieldwork, and the efforts of the first Africa Centre field team including D. Biyela, T. Ngubo, Z. Majozi, Z. Sithole, N. Zuma and S. Zungu. We thank F. Tanser, I. Timæus, N. Townsend, A-M. Vanneste, and anonymous reviewers for comments on previous drafts of this paper. We are also grateful to all household members who participated in the study and acknowledge their important contribution to the development of ACDIS.

References

Berry, S. 1993. *No Condition Is Permanent: The Social Dynamics of Agrarian Change in Sub-Saharan Africa*. Madison: University of Wisconsin Press.

Binka, F.N., Ngom, P., Phillips, J.F., Adazu, K. and MacLeod, B.B. 1999. "Assessing population dynamics in a rural African society: the Navrongo demographic surveillance system", *Journal of Biosocial Science* **31**(3):373–391.

Bledsoe, C. 1995. "Marginal members: children of previous unions in Mende households in Sierra Leone", in Greenhalgh, S. (ed.). *Situating fertility: anthropology and demographic inquiry*. New York: Cambridge University Press, pp. 130–153.

Bledsoe, C. and Pison, G. 1994. "Introduction", in Bledsoe, C. and Pison, G. (eds.). *Nuptiality in sub-Saharan Africa. Contemporary anthropological and demographic perspectives*. Oxford, England: Clarendon Press, pp. 1–22.

Booysen, F. 2003. *Out–migration in the context of the HIV/AIDS epidemic: evidence from the Free State Province*. CSSR Working paper No. 57. Cape Town: University of Cape Town. Available: http:/www.uct.ac.za/depts/cssr/pubs.html

Booysen, F.lR. and Arntz, T. 2003. "The methodology of HIV/AIDS impact studies: a review of current practices", *Social Science and Medicine* **56**:2391–2405.

Bray, R. 2002. *Missing links? An examination of contributions made by social surveys to our understanding of child well–being in South Africa*, CSSR Working paper No. 23. Cape Town: University of Cape Town. Available: http:/www.uct.ac.za/depts/cssr/pubs.html

Bray, R. 2003. "Who does the housework? An examination of South African children's working roles", *Social Dynamics* **29**(2):95–131.

Budlender, D. 2003. "The debate over household headship", *Social Dynamics* **29**(3): 48–70.

Budlender, D., Chobokoane, N. and Simelane, S. 2005."Marriage patterns in South Africa: methodological and substantive issues", *Southern African Journal of Demography* 9(1):1–25.

Burch, T.K. 1995. "Theories of household formation: progress and challenges", in Van Imhoff, E., Kuijsten, A., Hooimeijer, P. and Van Wissen, L. (eds.). *Household demography and household modelling*. New York: Plenum Press, pp. 85–107.

Caldwell, J. and Hill, A. 1988. "Recent developments using micro-approaches to demographic research", in Caldwell J. and Hull, V. (eds.). *Micro-approaches to demographic research*. London: Kegan Paul International.

Case, A. and Ardington, C. 2004. *ACDIS Monograph: Socioeconomic Factors*. Mtubatuba: Africa Centre for Health and Population Studies.

Cross, C. and Preston-Whyte, E. 1985. Manual for drawing household ideographs.

Curtis, B., Bradshaw, D. and Nojilana, B. 2002. *Socio-demographic Profile of the Magisterial District of Hlabisa – 1996 Census*. Durban: Medical Research Council, South Africa.

Das Gupta, M., Aaby, P., Garenne, M. and Pison, G. 1997. *Prospective community studies in developing countries*. International Studies in Demography. Oxford, England: Oxford University Press.

de Jong Gierveld, J. 1995. "The collection of survey data on household structures", in Van Imhoff, E., Kuijsten A., Hooimeijer, P. and Van Wissen, L. (eds.). *Household demography and household modelling*. New York: Plenum Press, pp. 137–162.

Demographic and Health Surveys. 1990. *Interviewer's manual for use with model 'A' questionnaire for high contraceptive prevalence countries*. DHS-II Basic documentation Number 3. Columbia, Maryland: Institute for Resource Development/Macro International, Inc.

Ferguson, J. 1999. *Expectations of modernity: myths and meanings of urban life on the Zambian copperbelt*. Berkeley & Los Angeles: University of California Press.

Fortes, M. 1966. "Introduction", in Goody, J. (ed.). *The development cycle in domestic groups*. Cambridge, England: Cambridge University Press, pp. 1–14.

Gage, A.J., Sommerfelt, A.E. and Piani, A.L. 1996. *Household structure, socio-economic level, and child health in sub-Saharan Africa.* DHS Analytical Reports, 1. Calverton, Maryland: Macro International Inc.

Garenne, M. 1997. "Demographic Surveillance System: concepts and methods developed in Niakhar (Senegal)", in *IUSSP general conference*. Beijing.

Gay, Judith. 1980. "Basotho Women's Options. A study of Marital Careers in Rural Lesotho." Cambridge University. Unpublished PhD dissertation.

Goody, E. 1982. *Parenthood and social reproduction*. Cambridge, England: Cambridge Studies in Social Anthropology.

Greenhalgh, S. 1996. "The social construction of population science: an intellectual, institutional, and political history of twentieth-century demography", *Comparative studies in society and history* 38(1):26–66.

Hill, C., Hosegood, V. and Gafos, M. 2005. Non-resident members of households in rural South Africa. Poster presented at International Union for the Scientific Study of Population. Tours, France. 18th–23rd July, 2005.

Hosegood, V. 1998. *The Hlabisa DSS pilot study – results from the household composition exercise*. Mtubatuba: Africa Centre for Population Studies and Reproductive Health, Population Studies Group.

Hosegood, V. and Timæus, I.M. 2005. "Household composition and dynamics in KwaZulu-Natal, South Africa: mirroring social reality in longitudinal data collection", in Van der Walle, E. (ed.). *African Households: An Exploration of Census Data*. New York: M.E. Sharpe Inc. pp 58–77.

Indepth Network. 2002. Chapter 1. "Core concepts of DSS", in *Population and health in developing countries. Volume 1. Population, Health, and Survival in INDEPTH Sites*. IDRC, Ottawa.

Izzard, W. 1985. "Migrants and mothers: case-studies from Botswana", *Journal of Southern African Studies* **11**(2):258–280.

Jones, S. 1993. *Assaulting Childhood. Children's Experiences of Migrancy and Hostel Life in South Africa.* Johannesburg: Witwatersrand University Press.

Jones, S. 1998. "Enacted marriages and fatherhood without jural paternity: signs of bilateral kinship among Xhosa in an Eastern Cape township", in *Association for Anthropology in Southern Africa*. University of Stellenbosch, South Africa.

Keilman, N. 1995. "Household concepts and household definitions in western Europe: different levels but similar trends in household development", in Van Imhoff, E., Kuijsten, A., Hooimeijer, P., and Van Wissen, L. (eds.). *Household Demography and Household Modelling*. New York: Plenum Press, pp. 111–135.

Kertzer, D. and Fricke, T. 1997. "Toward an anthropological demography", in Kertzer, D. and Fricke, T. (eds.). *Anthropological demography*. Chicago: University of Chicago Press, pp. 1–35.

Leliveld, A. 1997. "The effects of restrictive South African migrant labor policy on the survival of rural households in Southern Africa: a case study from rural Swaziland", *World Development* **25**(11):1839–1849.

Locoh, T. 1994. "Social change and marriage arrangements: new types of union in Lome, Togo", in Bledsoe, C. and Pison, G. (eds.). *Nuptiality in sub-Saharan Africa*. Oxford: Clarendon Press.

Lurie, M., Harrison, A., Wilkinson. D. and Karim, S.A. 1997. "Circular migration and sexual networking in rural KwaZulu/Natal: implications for the spread of HIV and other sexually transmitted diseases". *Health Transition Review* 7(3):17–27.

Marcus, T. 1998. *Sentinel Household Selection & Socio-economic Status Report*. CASE report no. 98/19. Pretoria: Community Agency for Social Enquiry.

Murray, C. 1978. "Migration, differentiation and the developmental cycle in Lesotho", in Van Binsbergen, W.M.J. and Meilink, H.A. (eds.). *Migration and the transformation of modern African society*, Volume 1. African Perspectives, pp. 127–144.

Murray, C. 1980. “Migrant labour and changing family structure in the rural periphery of Southern Africa”, *Journal of Southern African Studies* **6**(2):139–156.

Murray, C. 1981. *Families Divided: the Impact of Migration in Lesotho*. Cambridge: Cambridge University Press.

Niehaus, I. 1994. “Disharmonious spouses and harmonious siblings: conceptualizing household formation among urban residents of Qwaqwa”, *African Studies* **53**(1): 115–136.

Nunn, A.J., Wagner, H.U., Kamali, A., Kengeya-Kayondo, J.F. and Mulder, D.W. 1995. “Migration and HIV-1 seroprevalence in a rural Ugandan population”, *AIDS* **9**(5):503–506.

O’Laughlin, B. 1998. “Missing men? The debate over rural poverty and women-headed households in southern Africa”, *The Journal of Peasant Studies* **25**:1–48.

Ross, F. 2003. “Dependents and dependence: a case study of housing and heuristics in an informal settlement in the Western Cape”, *Social Dynamics* **29**(2):132–152.

Ross, F.C. 1996. “Diffusing domesticity: domestic fluidity in Die Bos”, *Social Dynamics* **22**(1): 55–71.

Russell, M. 2003a. “Are urban Black families nuclear? A comparative study of Black and White South African family norms”, *Social Dynamics* **29**(2):153–176.

Russell, M. 2003b. “Understanding black households: the problem”, *Social Dynamics* **29**(2):5–47.

Schapera, I. 1947. *Migrant labour and tribal life*. London, England: Oxford University Press.

Solarsh, G., Benzler, J., Hosegood, V., Tanser, F. and Vanneste, A. 2002. Chapter 18. “Hlabisa DSS”, in Indepth Network (ed.). *Population and health in developing countries. Volume 1. Population, Health, and Survival in INDEPTH Sites*. IDRC, Ottawa.

South Africa DHS. 1999. *South Africa demographic and health survey 1998*. MRC/Department of Health/Macro International Inc.

Spiegel, A. 1980. “Rural differentiation and the diffusion of migrant labour remittances in Lesotho”, in P. Mayer (ed.). *Black villagers in an industrial society. Anthropological perspectives on labour migration in South Africa.* Cape Town: Oxford University Press, pp. 109–168.

Spiegel, A. 1996. “Introduction: domestic fluidity in South Africa”, *Social Dynamics* **22**(1):5–6.

Spiegel, A., Watson, V. and Wilkinson, P. 1996. “Domestic diversity and fluidity among some African households in Greater Cape Town”, *Social Dynamics* **22**(1):7–30.

Spiegel, A.D. 1986. “The fluidity of household composition in Matatiele, Transkei: a methodological problem”, *African Studies* **45**(1):17–35.

Timæus, I.M. and Graham, W. 1989. “Labor circulation, marriage, and fertility in

Southern Africa", in Lesthaeghe, R.J. (ed.). *Reproduction and Social Organization in sub-Saharan Africa*. Los Angeles: University of California Press, pp. 365–400.

Townsend, N. 1997a. "Reproduction in anthropology and demography", in Kertzer, D. and Fricke, T. (eds.). *Anthropological Demography: Toward a New Synthesis*. Chicago: University of Chicago Press, pp. 96–114.

Townsend, N.W. 1997b. "Men, migration, and households in Botswana: an exploration of connections over time and space", *Journal of Southern African Studies* **23**(3): 405–420.

Van der Waal, C.S. 1996. "Rural children and residential instability in the Northern Province of South Africa", *Social Dynamics* **22**(1):31–54.

Watts, C. and Shrader, E. 1998. "How to do (or not to do) … The genogram: a new research tool to document patterns of decision-making, conflict and vulnerability within households", *Health Policy and Planning* **13**(4):459–464.

SUMMARY GUIDELINES FOR AUTHORS

(Full guidelines can be obtained from the editors.)

In the first instance, submissions to the Journal should conform to the standards of writing, analysis and presentation appropriate to an international peer-reviewed journal.

1. Submissions should be written in acceptable English of an academic quality, and saved in a word processing format compatible with Word 97 (or later). The editors prefer to receive submissions via email, but submission on 3.5" disk by mail to the address given at the front of this volume is also acceptable.
2. There is no specified upper limit to the length of articles that will be published, but submissions of more than 7 000 words are less likely to be accepted for publication.
3. Submission of a paper will be held to imply that it contains original unpublished work which is not being (nor has been) submitted for publication elsewhere.
4. Pages must be numbered consecutively with Arabic numerals. The first page should contain only the title, the author's name(s), affiliation(s) and an address for correspondence by both email and post. The second page should have only an abstract of approximately 150 words.
5. All manuscripts must be at least one-and-a-half spaced. The only exceptions permitted are footnotes and extended quotations, which may be single spaced.
6. Emphasis in the text, which is again discouraged, should be achieved through the use of italicisation, not underlining.
7. Paragraphs, other than the first after a heading, should be indented. No line breaks should be inserted between paragraphs, or between headings and the text.
8. Spelling and grammar must be set to "UK English". Particular attention should be paid to using the word 'data' as a plural noun (e.g. "The data show …", not "The data shows …".
9. Endnotes should not be used. Footnotes should be avoided unless their inclusion is essential to the text but not suitable for inclusion in the main body of the article.
10. Only those figures and tables deemed essential to the substance of the article should be included. Tables should be formatted in Word and numbered consecutively with Arabic numerals (e.g. Table 1; Table 2).
11. Figures should be formatted in Excel, and included as an insert in a Word document as a metafile. The underlying data should also be included separately to facilitate reproduction. The original graphics files of illustrations or maps embedded in word processing documents should also be supplied separately.
12. Equations should be inserted with Equation Editor. The use of standard definitions (e.g. nqx) is encouraged, while acronymic representation is strongly discouraged.
13. References must be included in the text, in parentheses, and indicate the author, year and – if a direct quote is used – the page number. The author's name may be omitted if the context is sufficiently clear. For example, "Lesthaeghe (1996) argues that …". Extended quotations (i.e. more than four lines) should be single spaced and indented from both sides. Authors should consult recent issues of the Journal for its house referencing style. An endnote style file is available from the editors on request.