

APPENDIX C

DATA QUALITY TABLES

The purpose of this Appendix is to provide the data user with a view of the general quality of the SADHS data. The tables in this appendix refer to possible *non-sampling* errors: digit preference, rounding or heaping on certain ages or dates; omission of events occurring further in the past; deliberate distortion of information by some interviewers in an attempt to lighten their workloads; non-cooperation of the respondent in providing information or refusal to be measured and weighted, etc. A description of the magnitude of such errors is provided in the following paragraphs.

Only 89 cases had missing information which accounted for 0.15 percent. The distribution of the de facto household population by single year of age is presented in Table C.1. The data show very little preference to report ages that end in zeros and fives (age “heaping” or digit preference) that is commonly found in countries where ages are not known well. There is some evidence of irregularities in the age distribution.

There is some evidence that interviewers “displaced” women age 15 and 49 years outside of the eligible range (15-49) presumably in order to avoid the need to interview them. For example, the number of women age 14 is substantially higher than the number age 15. For women, the number reported at age 14 (863) is more than that reported at age 13 (757) and 362 more than that at age 15 (863). At the other end of the range, the number of women age 49 is lower than the number age 50, implying that interviewers assigned an age 50 (or 51) to women in order to avoid interviewing them. A comparative study of DHS surveys noted some severe displacement out of the eligible age range (Rutstein and Bicego, 1990). The men also appear to have been “displaced” and there are a relatively high number age 14 (824) compared to age 15 (564).

The age distribution of eligible women from the de facto household population is compared with the age distribution of the sampled women interviewed. Table C.2 shows that the distributions are very similar and that the response rate was consistent across all ages. From Table C3, it can be seen that the response rate in the adult survey was higher for women (94 percent) than for men (85 percent). It was slightly higher for older men (90 percent) than for those under 65 years of age.

Information on the completeness of reporting selected important variables is provided in Table C.4. Overall, the percentage of cases with missing information is low. Month of birth was missing for 2.2 percent of births that occurred in the 15 years before the survey and, both month and year were recorded missing for less than 1 percent. Age at death was missing for 8.4 percent of non-surviving births which problematic. The size of child at birth had a sizeable proportion missing (21.8%). The missing information on the size of the child at birth is more likely to be attributed to the child not having been weighed at birth than interviewer negligence. A fairly high proportion of children had missing information on diarrhoea (5 percent).

According to Table C.5, the information on birth dating is of good quality: both month and year of birth were provided for 97 percent of all births and for 98 percent of surviving children. As expected, information on birth dates is more complete for children who were still living at the time of the survey than for those who had died. The sex ratio at birth was 103.4, which is very much what would be expected, but there was some variation from year to year.

There is very little evidence of transference of births out of 1994 to earlier years to avoid lengthier questionnaires. In fact, the ratio of births in 1994 to the average of the two adjoining years is 103 which shows slightly more births in 1994 than the average births of the two adjoining years.

Measurement of childhood deaths through retrospective household surveys often suffers from under reporting of deaths, in particular those deaths which occur very early in infancy. If early neonatal deaths are selectively under reported, the result would be an abnormally low ratio of deaths under seven days to all neonatal deaths and an abnormally low ratio of neonatal to infant mortality. Changes in these ratios over time can be examined to detect the hypothesis that under reporting of early deaths is more common for births that occurred longer before the survey.

Table C.6 shows the distribution of deaths under one month of age at death in days, while Table C.6 shows the distribution of deaths under two years of age by age of death in months. The data suggest that early infant deaths have *not* been severely under reported in the SADHS, since the percentage of early neonatal deaths during the neonatal period (next-to-last row in Table C.6) are reasonable. There is evidence of heaping on 7 days but this does not detract from the reasonable quality of the information. From Table C.6 it can be seen that the proportion of infant deaths that occur during the neonatal period is also reasonable and fairly stable at 45-50 percent.

The sibling history section of the women's questionnaire compiles a list of all of a respondent's brothers and sisters and collects information on the survival status of each of these siblings, the ages of living siblings and the ages at death and years since death of siblings who have died. These data can be used to measure all-cause adult mortality directly as well as maternal mortality.

Respondents provided information on 50,323 siblings, 6367 of whom were reported to have died (see Table C.8). Respondents knew the survival status of almost all of the siblings they reported and were able to report an age for 95.3 percent of their living siblings. However, they could only report both an age at death and the interval in years since the death for 66.8 percent of their dead siblings. While either an age at death or an interval since death was reported for slightly more than half of these dead siblings, neither item of information is available for 15.5 percent of them. The data on sisters and brothers suffer from about the same degree of incomplete reporting.

If reporting about siblings is accurate, one would expect their aggregate age distribution to resemble that of the respondents, which should be representative in turn of that of the population as a whole. Table C.9 provides these age distributions. As only women aged 15 to 49 were interviewed and they have siblings who fall outside this age range, the siblings should have a rather more dispersed age distribution than the respondents. Moreover, in a growing population the siblings should be slightly older on average than the respondents. The data conform to these patterns. Thus, Table C.9 provides no evidence of serious omissions of older brothers and sisters from the sibling histories or of major biases in the reporting of siblings' ages.

As most of the respondents' mothers have reached the end of their childbearing ages, the average size of respondents' sibships (i.e. the respondent herself plus all of her brothers and sisters) should be approximately equal to total fertility at the time when the respondents were born. Moreover, the ratio of brothers ever-born to sisters ever-born should be about the same as the sex ratio at birth. As can be seen from Table C.10, the sibling histories indicate a plausible downward trend in fertility in South Africa. Moreover, the sex ratio of the reported siblings of 1.01 does not differ significantly from a plausible sex ratio at birth for South Africa of 1.03 and is broadly the same for all cohorts of respondents. These statistics provide no evidence of poor quality reporting on siblings by older respondents.

In summary, the only important limitation of the sibling history data revealed by these simple checks is that the ages at death of about 23 percent of dead siblings had to be imputed.

Adjustment of DHS provincial estimates

The level of under-five mortality observed in SADHS was lower than that derived from the 1996 census (Udjo, 1998) and even lower than that from the 1993 Living Standards and Development Survey - (SALDRU 1994). This is likely to result from differences in the methodology used in the respective surveys - the census and the LSDS used indirect methods based on the proportion of children ever born who are still alive while SADHS collected a detailed pregnancy history enabling a direct estimate. The approach used in SADHS is usually considered to be more reliable. When the provincial estimates of under-five mortality in SADHS are compared with those in the census and those in the LSDS however, it is found that there is a good correlation between them in all the provinces excepting the Western Cape, Free State and North West where they appear to be too low in SADHS. This suggests that an adjustment for the underestimates for these three provinces would be appropriate.

An adjustment was made by using the underlying relationship between the child mortality observed in SADHS and the census results that were observed in the remaining provinces where there was good correlation. The child mortality in the three provinces could then be predicted from the level that was observed for that province in the census. The relationship between infant and child mortality observed in SADHS, excluding the three provinces could then be used to derive an adjusted infant mortality rate. The adjusted national and provincial rates are shown in Table C.11. Allowing for the underestimates in these three provinces the adjustment suggests that the national child mortality rates should be adjusted by a factor of 6.1%. While the adjusted rates should provide good estimates of child mortality for each province, reasons for the underestimates in these three provinces need to be explored.

The census and the Living Standards and Development Survey (LSDS) estimates of under five mortality are derived using a method developed by Brass (Brass 1964). The estimates used from the census correspond to the reference periods 1994.1 and 1992.3 using the data of women aged 20-24 and 25-29 respectively. The corresponding reference period from the LSDS is 1992.3 using the data of women aged 20-24. Table C.12 compares these to the SADHS rates which are based on the period 10 years before the survey and centres on the year 1993.

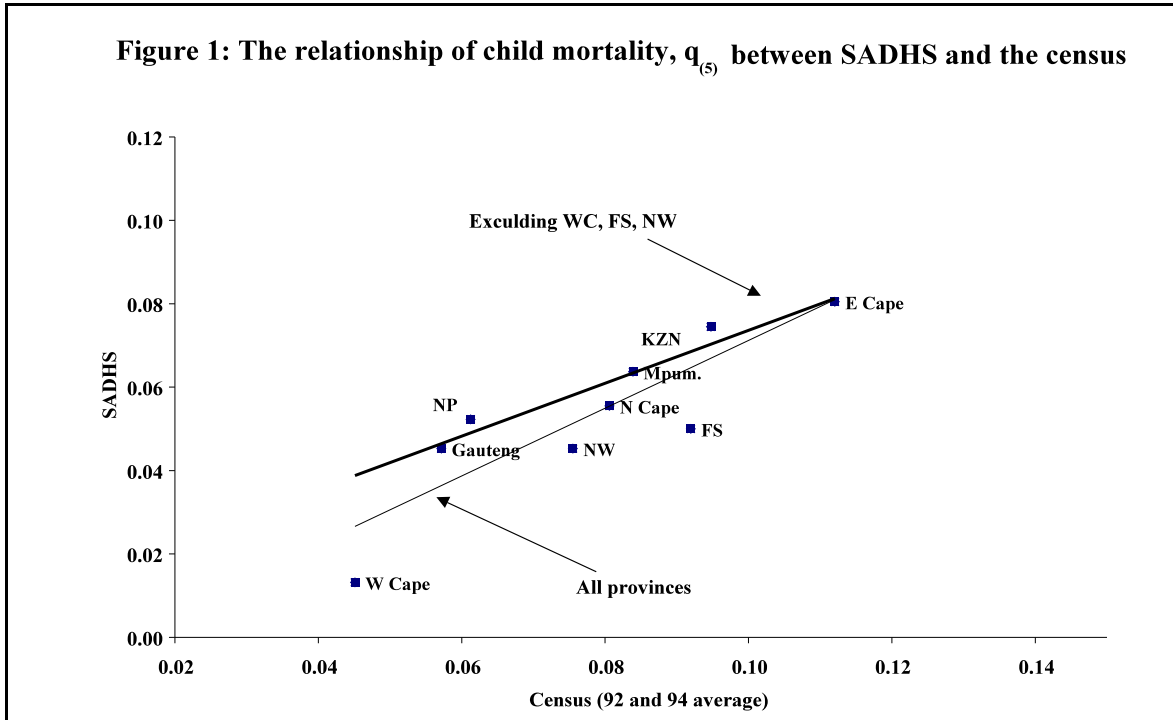
As can be seen from Figure 1, the relationship between the provincial $q_{(5)}$ estimates from the census and those from the SADHS show that there is a very good correlation between them excepting for the Western Cape, Free State and North West provinces where the mortality observed in SADHS appears to be lower than what would be expected. The correlation between the two sets of estimates show that the census 1992 estimate has the most agreement with the SADHS. An average of the 1992 and 1994 values gives a higher correlation. This average has been used to estimate the relationship between the SADHS child mortality and the census level. The correlation coefficients for the SADHS and the census for 1992, 1994, and an average of 1992 and 1994 are, 0.917, 0.925 and 0.93 respectively.

The relationship between $q_{(5)}$ estimates from the census and $q_{(5)}$ from the SADHS are shown in the graph and a line is fitted for the 6 provinces with reliable estimates which predicts $q_{(5)}$ for the 3 underestimated provinces using the equation :

$$q_{(5)} = 0.6343 * \text{census } q_{(5)} + 0.0102$$

Infant mortality is derived using the adjusted $q_{(5)}$ estimates and the equation:

$$q_{(1)} = 0.6920 * \text{adjusted } q_{(5)} + 0.0031$$



References

Brass W 1964 cited in Manual X. Indirect Techniques for Demographic Estimation. Department of International and Social Affairs. Population Studies Report No.81. United Nations, New York 1983.

SALDRU 1994. South Africans rich and poor: Baseline household statistics. Cape Town: South Africa Labour Development Research Unit, University of Cape Town.

Udjo E 1997. Additional evidence regarding fertility and mortality trends in South Africa and implications for population projections. Statistics South Africa, Pretoria.

Table C.1 Household age distribution

Single-year age distribution of the de facto household population by sex (weighted), South Africa 1998

Age	Male		Female		Age	Male		Female	
	Number	Percent-age	Number	Percent-age		Number	Percent-age	Number	Percent-age
0	552	2.3	558	2.0	37	254	1.0	351	1.3
1	517	2.1	550	2.0	38	293	1.2	341	1.2
2	542	2.2	536	1.9	39	253	1.0	329	1.2
3	562	2.3	541	2.0	40	291	1.2	337	1.2
4	588	2.4	622	2.3	41	190	0.8	255	0.9
5	613	2.5	588	2.1	42	266	1.1	289	1.0
6	739	3.1	703	2.5	43	247	1.0	259	0.9
7	744	3.1	717	2.6	44	176	0.7	234	0.8
8	644	2.7	676	2.4	45	231	1.0	253	0.9
9	621	2.6	668	2.4	46	167	0.7	215	0.8
10	697	2.9	688	2.5	47	176	0.7	206	0.7
11	631	2.6	662	2.4	48	168	0.7	207	0.7
12	764	3.2	753	2.7	49	186	0.8	157	0.6
13	701	2.9	757	2.7	50	212	0.9	441	1.6
14	824	3.4	863	3.1	51	127	0.5	252	0.9
15	564	2.3	501	1.8	52	160	0.7	282	1.0
16	571	2.4	482	1.7	53	159	0.7	220	0.8
17	582	2.4	464	1.7	54	154	0.6	196	0.7
18	567	2.3	521	1.9	55	146	0.6	202	0.7
19	481	2.0	428	1.5	56	132	0.5	196	0.7
20	459	1.9	442	1.6	57	110	0.5	204	0.7
21	444	1.8	475	1.7	58	152	0.6	188	0.7
22	385	1.6	474	1.7	59	126	0.5	197	0.7
23	377	1.6	449	1.6	60	141	0.6	248	0.9
24	397	1.6	402	1.5	61	103	0.4	153	0.6
25	372	1.5	463	1.7	62	104	0.4	198	0.7
26	294	1.2	334	1.2	63	102	0.4	145	0.5
27	335	1.4	410	1.5	64	106	0.4	188	0.7
28	314	1.3	401	1.5	65	149	0.6	194	0.7
29	298	1.2	358	1.3	66	79	0.3	100	0.4
30	336	1.4	376	1.4	67	98	0.4	139	0.5
31	247	1.0	319	1.2	68	98	0.4	150	0.5
32	278	1.1	377	1.4	69	84	0.3	124	0.4
33	234	1.0	358	1.3	70+	618	2.6	1,146	4.1
34	269	1.1	359	1.3	Don't know				
35	309	1.3	382	1.4	missing	51	0.2	38	0.1
36	252	1.0	345	1.2					
					Total	24,215	100.0	27,643	100.0

Note: The de-facto population includes all residents and non residents (visitors) who slept in the household the night before the interview.

Table C.2 Age distribution of eligible and interviewed women

Percent distribution by age of the de facto household population of women age 10-54, and of interviewed women age 15-49, and percentage of eligible women who were interviewed (weighted), South Africa 1998

Age	Household		Women interviewed		
	Number	Percent-age	Number	Percent-age	Percentage interviewed
10-14	3,724	NA	NA	NA	NA
15-19	2,397	19.1	2,275	19.3	94.9
20-24	2,243	17.9	2,099	17.8	93.6
25-29	1,966	15.7	1,855	15.8	94.3
30-34	1,790	14.3	1,680	14.3	93.8
25-39	1,747	13.9	1,623	13.8	92.9
40-44	1,374	10.9	1,279	10.9	93.1
45-49	1,039	8.3	965	8.2	92.9
50-54	1,392	NA	NA	NA	NA
15-49	12,557	NA	11,775	NA	93.8

NA= Not applicable

Note: The de facto population includes all residents and non-residents who slept in the household the night before interview.

Table C.3 Age distribution of eligible and interviewed adults

Percent distribution by age of the de facto household population of adults age 15 and above and of the interviewed adults age 15 and above, and percentage of eligible adults who were interviewed (weighted), South Africa 1998

Age	Adult men 15+					Adult women 15+				
	Household		Men interviewed			Household		Women interviewed		
	Number	Percentage	Number	Percentage	Percentage interviewed	Number	Percentage	Number	Percentage	Percentage interviewed
15 - 24	2,056	31.4	1,805	32.3	87.8	2,157	25.3	2,041	25.7	94.6
25 - 34	1,310	20.0	1,095	19.6	83.6	1,824	21.4	1,690	21.3	92.6
35 - 44	1,205	18.4	1,992	17.8	82.3	1,530	18.0	1,408	17.7	92.0
45 - 54	841	12.8	687	12.3	81.7	1,155	13.6	1,087	13.6	94.1
55 - 64	580	8.9	506	9.1	87.2	958	11.2	905	11.3	94.5
65+	555	8.5	502	9.0	90.4	896	10.5	853	10.7	95.2
15+	6,547	100.0	5,587	100.0	85.3	8,520	100.0	7,984	100.0	93.7

Table C.4 Completeness of reporting

Percentage of observations missing information for selected demographic and health questions, South Africa 1998

Subject	Reference group	Percentage with missing information	Number
Birth Date	Last 15 years		
Month only		2.16	15,289
Month and year		0.72	15,289
Age at death	Death to births in last 15 years	8.35	906
Age/date at first union ¹	Ever-married women	2.51	6,070
Respondent's education	All women 15-49	0.00	11,735
Child's size at birth	Births in last 1-59 months	21.77	4,345
Diarrhoea in last 2 weeks	Living children age 1-59 months	4.98	4,740

¹Both year and age missing

Table C.5 Births by calendar year

Distribution of births by calendar year of birth since birth for living (L), dead (D), and all (A) children, according to reporting completeness, sex ratio at birth, and ratio of births by calendar year, South Africa 1998

Year	Total number of births			Percentage with complete birth date ¹			Sex ratio at birth ²			Calendar ratio ³			Number of male births			Number of female births		
	L	D	A	L	D	A	L	D	A	L	D	A	L	D	A	L	D	A
98	272	5	276	100.0	100.0	100.0	111.1	NA	114.9	-	-	-	143	5	148	129	0	129
97	1,014	61	1,075	98.9	90.3	98.4	94.3	146.1	96.6	158.0	210.1	160.2	492	36	528	522	25	547
96	1,012	53	1,065	99.1	89.8	98.7	97.2	157.6	99.6	107.0	90.3	106.0	499	32	531	513	21	534
95	878	57	935	98.0	99.1	98.0	95.5	151.4	98.2	90.5	114.1	91.7	429	34	463	449	22	472
94	928	46	974	98.0	97.8	98.0	110.6	148.6	112.1	103.9	86.8	103.0	487	28	515	441	19	459
93	907	50	957	98.9	83.3	98.1	104.5	82.0	103.2	89.0	78.8	88.4	464	22	486	444	27	471
92	1,112	80	1,192	97.7	89.8	97.2	93.4	138.9	95.9	108.9	170.1	111.6	537	47	584	575	34	608
91	1,134	45	1,179	98.4	78.0	97.6	106.9	77.7	105.6	106.9	69.4	104.8	586	19	605	548	25	573
90	1,010	48	1,058	98.0	88.4	97.6	96.4	139.3	98.0	96.7	89.9	96.4	496	28	524	514	20	534
89	954	63	1,017	97.1	84.6	96.3	88.3	190.2	92.4	-	-	-	447	41	488	507	22	528
94-98	4,104	221	4,325	98.6	94.2	98.4	99.8	156.4	102.1	-	-	-	2,051	135	2,185	2,054	86	2,140
89-93	5,117	285	5,403	98.0	85.4	97.3	97.8	123.4	99.0	-	-	-	2,530	158	2,687	2,588	128	2,715
84-88	4,485	355	4,840	97.8	91.0	97.3	99.5	139.1	101.9	-	-	-	2,237	206	2,443	2,248	148	2,396
79-83	3,641	358	3,999	97.5	83.1	96.2	111.9	121.6	112.7	-	-	-	1,923	196	2,119	1,718	162	1,880
< 79	3,599	589	4,189	97.1	82.9	95.1	101.2	118.4	103.5	-	-	-	1,811	320	2,130	1,789	270	2,059
All	20,947	1,808	22,756	97.8	86.3	96.9	101.5	127.9	103.4	-	-	-	10,551	1,015	11,566	10,396	794	11,190

NA = Not applicable

¹Both year and month of birth given

² $(B_m/B_f)*100$, where B_m and B_f are the numbers of male and female births, respectively

³ $[2B_x/(B_{x-1}+B_{x+1})]*100$, where B_x is the number births in calendar year x

Table C.6 Reporting of age at death in days

Distribution of reported deaths under 1 month of age by age at death in days and the percentage of neonatal deaths reported to occur at ages 0-6 days, for five year periods preceding the survey, South Africa 1998

Age at death (in days)	Number of years preceding the survey				Total 0-19
	0-4	5-9	10-14	15-19	
<1	12	21	17	16	67
1	28	19	33	26	105
2	14	11	11	10	45
3	4	7	7	6	24
4	4	1	4	3	11
5	2	6	3	0	10
6	2	5	1	0	9
7	7	4	6	7	25
8	0	0	3	3	6
9	4	0	0	3	7
10	6	1	4	1	13
12	1	0	0	0	1
13	0	0	1	0	1
14	7	8	3	5	23
15	0	1	0	4	5
16	0	0	0	0	1
18	0	1	0	1	2
19	0	1	0	0	1
20	2	1	0	0	3
21	2	2	4	1	9
22	0	0	2	0	3
25	1	0	0	0	1
30	0	1	6	0	7
Total 0-30	96	91	104	87	378
Missing	1	1	0	2	4
Percent early neonatal ¹	68.4	76.8	72.2	69.5	71.7

¹Percent early neonatal: 0-6 days/0-30 days

Table C.7 Reporting of age at death in months

Distribution of reported deaths under 2 years of age by age at death in months and the percentage of infant deaths reported to occur at ages under one month, for five-year periods preceding the survey, South Africa 1998

Age at death (in months)	Number of years preceding the survey				Total 0-19
	0-4	5-9	10-14	15-19	
< 1 ^a	97	92	104	89	382
1	18	16	14	17	64
2	16	17	15	7	55
3	10	10	23	20	63
4	13	6	10	8	37
5	10	5	4	9	29
6	6	13	16	14	49
7	6	1	9	8	23
8	11	6	5	12	34
9	8	12	14	13	46
10	4	5	3	2	14
11	4	5	6	4	18
12	15	20	24	18	77
13	0	1	1	0	2
14	2	1	0	5	8
15	2	1	3	2	8
16	0	0	2	1	4
17	0	1	3	1	5
18	0	0	1	7	8
19	0	0	0	0	1
20	0	0	0	0	1
1 Year	6	12	9	9	36
Total 0-11 ^b	201	187	224	202	814
Percent Neonatal ^c	48.1	49.0	46.7	44.2	46

^a Includes deaths under 1 month reported in days.

^b Includes cases for which age at death (in exact months) is not known.

^c Deaths under 1 month divided by deaths under 1 year multiplied by 100.

Table C.8 Data on siblings: completeness of the reported data

Number of siblings reported by survey respondents and completeness of the reported data on age, age at death (AD), and years since death (YSD), South Africa 1998

Sibling status and completeness of reporting	Sisters		Brothers		Total	
	Number	Percent	Number	Percent	Number	Percent
All siblings	25,053	100.0	25,270	100.0	50,323	100.0
Living	22,514	89.9	21,319	84.4	43,834	87.1
Dead	2,497	10.0	3,870	15.3	6,367	12.7
Missing survival status	41	0.2	81	0.3	122	0.2
Living siblings	22,514	100.0	21,319	100.0	43,834	100.0
Age reported	21,467	95.4	20,301	95.2	41,769	95.3
Age missing	1,047	4.7	1,018	4.8	2,065	4.7
Dead siblings	2,497	100.0	3,870	100.0	6,367	100.0
AD and YSD reported	1,636	65.5	2,619	67.7	4,255	66.8
Only AD missing	148	5.9	302	7.8	450	7.1
Only YSD missing	306	12.3	368	9.5	674	10.6
AD and YSD missing	408	16.3	580	15.0	988	15.5

* AD=age at death, YSD=years since death

Table C.9 Percent distribution of respondents and their siblings by birth cohort, South Africa, 1998

Year of birth	Repondents	Siblings
Before 1945	0.0	4.1
1945-49	2.1	5.3
1950-49	9.6	7.4
1955-59	12.1	11.6
1960-64	14.3	13.8
1965-69	14.4	14.6
1970-74	16.4	13.9
1975 or later	31.1	29.1
All years	100.0	100.0
Median year of birth	1969	1967
Number of cases	11,735	50,285

Table C.10 Mean sibship size and sex ratio at birth of siblings by respondents' birth cohort, South Africa, 1998

Year of birth	Mean sibship size	Sex ratio at birth
1940s	6.3	100.9
1950-54	6.0	94.3
1955-59	5.7	96.3
1960-64	5.6	103.1
1965-69	5.5	103.1
1970-74	5.2	98.3
1975-79	4.9	105.9
1980-84	4.5	104.1
All years	5.3	100.9

Table C.11 Adjusted provincial and national estimates of infant and under-five mortality rates observed in the 10 year period preceding the survey

Province	Adjusted Infant mortality rate per 1000 births	Adjusted under-five mortality rate per 1000 births
Western Cape*	30.0	38.8
Eastern Cape	61.2	80.5
Northern Cape	41.8	55.5
Free State*	50.5	68.5
KwaZulu Natal	52.1	74.5
North West*	43.3	58.1
Gauteng	36.3	45.3
Mpumalanga	47.3	63.7
Northern	37.2	52.3
South Africa	47.8	60.4

*These provinces have been adjusted on the basis of the relationship between SADHS and the 1996 census data observed in the remaining provinces.

Table C.12 Under-five mortality rates derived from the 1996 census, LSDS and the SADHS

Province	Census 1992.3	Census 1994.1	LSDS	SADHS
Western Cape*	43.8	46.4	113.9	13.2
Eastern Cape	110.2	113.9	157.5	80.5
Northern Cape	80.4	80.8	97.5	55.5
Free State*	91.2	92.7	181.3	50.0
KwaZulu Natal	93.2	96.3	113.2	74.5
North West*	73.4	77.6	130.6	45.3
Gauteng	54.1	60.2	66.9	45.3
Mpumalanga	86.7	81.1	104.4	63.7
Northern	60.1	62.4	79.6	52.3

*Child mortality in these provinces is considered to be under reported in SADHS when compared with the LSDS and the census.