

APPENDIX C

DATA QUALITY TABLES

APPENDIX C

DATA QUALITY TABLES

The purpose of this Appendix is to provide the data user with an initial view of the general quality of the ZDHS data. Appendix B is concerned with sampling errors and their effects on the survey results. 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 farther 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 weighed, etc. A description of the magnitude of such non-sampling errors is provided in the following paragraphs.

The distribution of the de facto household population by single year of age is presented in Table C.1 (see also Figure 2.1). The data show 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, e.g., relatively higher numbers among men at ages 9, 12, and 27 and among women at ages 13, 19, 23 and 27. However, it is difficult to find any pattern to these results and they may be due to random errors. The irregularities appear to be somewhat worse among women than among men.

There is also some evidence that interviewers "displaced" women outside of the eligible age range (15-49), presumably in order to avoid the need to interview them. For example, while the number of men age 15 is substantially higher than the number age 14 and 13 (perhaps due to the tendency mentioned above to round ages to the nearest age ending in zero or five), the number of women age 15 is substantially lower than the number age 14 or 13. At the other end of the range, the number of men age 49 exceeds the number age 50, while the converse is true for women, implying that interviewers assigned an age of 50 (or 51) to women whose ages might not have been known with certainty, in order to avoid interviewing them. In any case, this displacement out of the eligible age range is much less severe than in many other DHS surveys (Rutstein and Bicego, 1990).

Table C.2 shows that response rates vary little according to age of respondents. The five-year age distribution of respondents shows a larger proportion than expected in age group 15-19 (28 percent) and a smaller proportion in the older age groups, 35-49 (see also Chapter 2, Section 2.3). The fact that men enumerated in the household have a similar distribution at age groups 15-19 and 20-24 as the women (see Table C.1) indicates that the large proportion of women age 15-19 is not sex selective and might be real.

Information on the completeness of reporting selected important variables is provided in Table C.3. Overall, the percentage of cases with missing information is extraordinarily low. Month of birth was missing for two percent of births that occurred in the 15 years before the survey and both month and year were missing for less than half of one percent. Age at death was missing for an infinitesimal proportion of non-surviving births. Only for the anthropometric measurements is there a sizeable proportion for which data are missing; about 6 percent of children under five were not measured. This level of nonresponse is still very acceptable in comparison with rates from other DHS surveys. The main reason for not measuring children was that the child was not present, either because he/she did not live with the mother or because he/she was not home. Very few mothers refused to let their children be measured.

Table C.1 Household age distribution

Single-year age distribution of the de facto household population by sex (weighted), Zambia 1992

Age	Males		Females		Age	Males		Females	
	Number	Percent	Number	Percent		Number	Percent	Number	Percent
<1	702	4.2	730	4.2	36	148	0.9	136	0.8
1	596	3.6	572	3.3	37	156	0.9	131	0.8
2	607	3.6	618	3.6	38	139	0.8	127	0.7
3	521	3.1	527	3.1	39	134	0.8	121	0.7
4	517	3.1	528	3.1	40	111	0.7	106	0.6
5	520	3.1	582	3.4	41	100	0.6	76	0.4
6	566	3.4	559	3.2	42	118	0.7	119	0.7
7	492	3.0	562	3.3	43	108	0.6	129	0.7
8	468	2.8	553	3.2	44	94	0.6	104	0.6
9	554	3.3	506	2.9	45	119	0.7	93	0.5
10	469	2.8	563	3.3	46	95	0.6	83	0.5
11	439	2.6	443	2.6	47	65	0.4	75	0.4
12	507	3.0	502	2.9	48	80	0.5	83	0.5
13	396	2.4	523	3.0	49	104	0.6	76	0.4
14	426	2.6	495	2.9	50	98	0.6	115	0.7
15	454	2.7	427	2.5	51	83	0.5	113	0.7
16	435	2.6	444	2.6	52	91	0.5	126	0.7
17	420	2.5	422	2.4	53	66	0.4	94	0.5
18	384	2.3	401	2.3	54	58	0.3	70	0.4
19	381	2.3	430	2.5	55	70	0.4	104	0.6
20	342	2.1	309	1.8	56	84	0.5	69	0.4
21	319	1.9	315	1.8	57	55	0.3	54	0.3
22	289	1.7	320	1.9	58	48	0.3	62	0.4
23	274	1.6	359	2.1	59	70	0.4	64	0.4
24	235	1.4	228	1.3	60	97	0.6	78	0.5
25	259	1.6	238	1.4	61	46	0.3	51	0.3
26	196	1.2	241	1.4	62	58	0.3	35	0.2
27	283	1.7	307	1.8	63	42	0.3	42	0.2
28	248	1.5	226	1.3	64	46	0.3	41	0.2
29	177	1.1	229	1.3	65	44	0.3	49	0.3
30	202	1.2	202	1.2	66	31	0.2	17	0.1
31	168	1.0	196	1.1	67	40	0.2	29	0.2
32	218	1.3	197	1.1	68	34	0.2	31	0.2
33	166	1.0	201	1.2	69	29	0.2	29	0.2
34	145	0.9	167	1.0	70+	310	1.9	231	1.3
35	191	1.1	172	1.0	Don't know, missing	21	0.1	3	0.0
					Total	16662	100.0	17261	100.0

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

Table C.2 Age distribution of eligible and interviewed women

Five-year age distribution of the de facto household population of women aged 10-54, five year age distribution of interviewed women aged 15-49, and percentage of eligible women who were interviewed (weighted), Zambia 1992

Age	Household population of women		Interviewed women		Percent interviewed (weighted)
	Number	Percent	Number	Percent	
10-14	2526	NA	NA	NA	NA
15-19	2123	28.3	1984	28.1	93.4
20-24	1531	20.4	1441	20.4	94.1
25-29	1241	16.6	1179	16.7	95.0
30-34	964	12.9	915	13.0	94.9
35-39	687	9.2	656	9.3	95.6
40-44	534	7.1	505	7.2	94.6
45-49	410	5.5	380	5.4	92.6
50-54	518	NA	NA	NA	NA
15-49	7490	100.0	7060	100.0	94.3

Note: The de facto population includes all residents and nonresidents who slept in the household the night before interview.
NA = Not applicable

Table C.3 Completeness of reporting

Percentage of observations missing information for selected demographic and health questions (weighted), Zambia 1992

Subject	Reference group	Percentage missing information	Number of cases
Birth date	Births in last 15 years		
Month only		1.8	15825
Month and year		0.1	15825
Age at death	Deaths to births in last 15 years	0.1	2535
Age/date at first union ¹	Ever-married women	1.3	5269
Respondent's education	All women	0.1	7060
Child's size at birth	Births in last 59 months	0.2	3171
Anthropometry²	Living children age 0-59 months		
Height		5.7	5393
Weight		5.5	5393
Height and weight missing		6.0	5393
Diarrhoea in last 2 weeks	Living children age 0-59 months	2.2	5393

¹Both year and age missing

²Child not measured

Table C.4 Births by calendar year since birth

Distribution of births by calendar years since birth for living (L), dead (D), and all (T) children, according to reporting completeness, sex ratio at birth, and ratio of births by calendar year, Zambia 1992

Year	Number of births			Percentage with complete birth date ¹			Sex ratio at birth ²			Calendar ratio ³			Male			Female		
	L	D	T	L	D	T	L	D	T	L	D	T	L	D	T	L	D	T
92	284	17	301	99.6	100.0	99.6	100.2	100.3	100.2	NA	NA	NA	142	8	151	142	8	150
91	1297	162	1459	99.8	93.7	99.1	101.0	113.6	102.4	186.2	155.8	182.3	652	86	738	645	76	721
90	1108	191	1300	99.2	97.7	99.0	97.5	97.2	97.5	93.2	111.1	95.5	547	94	641	561	97	658
89	1082	182	1264	99.2	93.4	98.3	101.1	124.4	104.2	107.2	97.2	105.6	544	101	645	538	81	619
88	911	184	1094	99.2	98.1	99.0	96.3	113.7	99.0	91.6	97.2	92.5	447	98	544	464	86	550
87	907	196	1102	99.6	95.0	98.8	93.9	125.9	98.9	95.7	93.8	95.4	439	109	548	468	87	554
86	984	233	1217	99.1	95.2	98.4	92.3	102.0	94.1	108.5	117.1	110.0	472	118	590	512	115	627
85	908	203	1111	98.9	96.9	98.5	91.1	100.9	92.8	98.5	96.0	98.0	433	102	535	475	101	576
84	860	189	1049	98.6	93.5	97.7	80.2	114.9	85.6	99.7	101.7	100.1	383	101	484	477	88	565
83	817	169	985	98.3	94.4	97.6	105.4	102.2	104.9	NA	NA	NA	419	85	504	398	84	481
88-92	4682	736	5418	99.4	95.9	98.9	99.2	111.2	100.8	NA	NA	NA	2332	388	2719	2350	348	2698
83-87	4475	990	5465	98.9	95.0	98.2	92.1	108.6	94.9	NA	NA	NA	2146	515	2661	2329	474	2804
78-82	3649	705	4354	98.5	92.6	97.5	97.2	105.3	98.5	NA	NA	NA	1799	362	2160	1850	343	2194
73-77	2777	561	3338	97.6	90.1	96.3	101.9	86.2	99.1	NA	NA	NA	1402	260	1661	1376	301	1677
<73	2537	809	3346	95.2	86.9	93.2	101.7	96.6	100.4	NA	NA	NA	1279	398	1677	1258	412	1669
All	18119	3801	21920	98.2	92.3	97.2	97.8	102.3	98.5	NA	NA	NA	8957	1922	10878	9163	1879	11042

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 of births in calendar year x

According to Table C.4, the information on birth dating is good: both month and year of birth were provided for 97 percent of all births and for 99 percent of births in the five years before the survey. 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. Still, both month and year of birth were provided for 92 percent of non-surviving children. Sex ratios are somewhat on the low side; the expected value would be 102 to 103, while those from the ZDHS are often less than 100. This indicates some possible undercounting of male births, especially those that occurred earlier in time.

The data in Table C.4 also indicate that there was transference of births out of 1987 to earlier years, particularly to 1986. The ratio of births in 1987 to the average of the two adjoining years is 95. This is almost surely not accidental, but rather represents the deliberate attempt by some interviewers to lighten their workloads, since several sections of the ZDHS questionnaire are applicable to only those children born since January 1987. This transference of births across the five-year cutoff point has been noted in many other DHS surveys (Arnold, 1990); in fact, the level of transference is lower in Zambia than in most of the sub-Saharan countries covered (e.g., Botswana 93; Burundi 83; Ghana 101; Liberia 71; Mali 84; Nigeria 77; Senegal 89; Togo 81; Uganda 96; Zimbabwe 97).

Table C.5 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 of birth preceding the survey, Zambia 1992

Age at death (in days)	Number of years preceding the survey				Total 0-19
	0-4	5-9	10-14	15-19	
<1	63	41	12	21	138
1	40	24	17	5	85
2	30	21	10	12	74
3	15	11	9	7	42
4	16	9	9	2	36
5	12	7	6	2	28
6	2	3	2	2	9
7	31	40	30	19	120
8	1	2	3	0	5
9	1	0	2	0	3
10	2	0	1	2	5
11	0	0	1	0	1
12	2	1	1	1	6
13	1	0	0	2	3
14	24	23	17	4	67
15	0	1	1	0	2
16	1	0	0	1	2
17	0	1	1	0	2
18	2	0	0	0	2
19	0	1	0	0	1
20	0	1	0	0	1
21	16	10	7	7	40
22	0	1	0	0	1
23	0	0	0	1	1
25	2	1	0	0	3
27	1	0	0	0	1
28	2	1	1	1	5
29	1	0	0	0	1
30	0	2	3	3	8
31+	1	0	0	0	1
Total 0-30	264	202	133	94	693
Percent early neonatal ¹	67.5	57.5	48.7	55.2	59.3

¹(0-6 days/0-30 days) * 100

Measurement of childhood deaths through retrospective household surveys often suffers from underreporting of deaths, in particular those deaths which occur very early in infancy. If early neonatal deaths are selectively underreported, 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 underreporting of early infant deaths is more common for births that occurred longer before the survey.

Table C.5 shows the distribution of deaths under one month of age by age at death in days, while Table C.6 shows the distribution of deaths under two years of age by age at death in months. The data suggest that early infant deaths have *not* been severely underreported in the ZDHS, since the percentage of neonatal deaths occurring in the first 6 days (next-to-last row in Table C.5) and the percentage of infant deaths

occurring during the neonatal period (next-to-last row in Table C.6) are reasonable. The former proportions increase over time, implying that some early infant deaths were not reported in the earlier periods; however, much, if not all, of this pattern can be attributed to heaping on 7 days at death, which is more severe for the earlier periods. As mentioned in Chapter 7, there was very little heaping on age at death of 12 months.

Table C.6 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 of birth preceding the survey, Zambia 1992

Age at death (in months)	Number of years preceding the survey				Total 0-19
	0-4	5-9	10-14	15-19	
<1 ^a	264	202	133	94	693
1	41	27	21	18	106
2	37	30	28	12	107
3	44	29	16	12	101
4	38	25	19	11	92
5	40	27	16	5	88
6	35	41	28	17	121
7	32	19	10	10	70
8	25	23	21	12	82
9	31	25	19	10	85
10	19	13	11	10	53
11	18	19	8	1	47
12	25	26	22	15	87
13	24	21	21	8	74
14	19	38	11	17	84
15	19	17	12	10	58
16	18	14	5	5	43
17	12	20	5	5	42
18	14	47	22	20	104
19	14	9	8	8	38
20	17	20	9	2	47
21	11	11	7	1	30
22	1	4	0	0	6
23	4	7	1	1	13
24+	5	1	7	2	16
1 year	3	12	7	5	28
Total 0-11	623	480	330	213	1646
Percent neonatal ^b	42.4	42.1	40.2	44.1	42.1

^aIncludes deaths under 1 month reported in days

^b(Under 1 month/under 1 year) * 100

APPENDIX D

PERSONS INVOLVED IN THE ZAMBIA DEMOGRAPHIC AND HEALTH SURVEY

