

KWAZULU-NATAL INCOME DYNAMICS STUDY (KIDS)
THIRD WAVE, 2004
(PUBLIC RELEASE VERSION 1.1)

Notes on the Household Roster

In KIDS 2004, the questionnaires pre-listed all individuals who had been a member of the household either in 1993 and 1998 and inquired about them. Some of these individuals no longer have any connection with panel households. Thus, roster records exist for individuals who are not household members. In addition, only limited information is available on non-resident members.

Version 1.1 of the 2004 household roster contains 11813 records on 11311 unique individuals, including 8496 records on 8277 unique resident individuals. The roster contains 2 records for 431 people, 3 records for 34 people and 4 records for one person. Among these 466 people with multiple records, 240 are no longer a member of any household, 22 belong to one household, 202 to two households and 2 to three households.

Of the 204 members with a membership record for more than one household, all but 8 spent 15+ days of the last month in a study household: 171 live in one of the households that they belong to, and 25 are reported to reside in two different households.

Group of records	Version 1.1 (records)	Unique people
Ex-members reported by core HHs – alive	1782	1497
Ex-members reported by core HHs – survival not known	67	62
Ex-members reported as decision-makers by <i>k</i> and <i>n</i> HHs	9	3
Deaths since 1998 wave	602	602
Living members from previous waves	5705	5499
Living new members	3648	3648
Total	11813	11311

The roster records are divided between the different types of household as follows:

Type of person	Extinct Core	Core	Next Generation	Foster Child	Total
Dead member	69	533	0	0	602
New member	0	1654	1054	940	3648
Old member	0	4494	809	402	5705
Non-member	0	1849	9	0	1858
Total	69	8530	1872	1342	11813
Resident member	0	5355	1703	1244	8302
N of households	49	865	319	193	1426

The following derived variables have been added to the roster files:

- 1) A new age variable, *age_dob*, has been calculated using date of interview and date of birth, if it was available, and the 1998 age and dates of interview if the 2004 data were missing or impossible (i.e. after the 1998 wave for someone reported in 1998).
- 2) Total expenditure before death, *tot_c_edt*, was calculated for the death records.

Longitudinal linkage of the roster data

It should be possible to link everyone in the 2004 roster who was interviewed in the 1993 and 1998 waves (i.e. *pcode*<100) to the 1998 data. However, merging with a 7-digit *hhid* fails for 4 people who moved between two 1998 households within a 1993 dynasty between 1998 and 2004. These 4 individuals need to be linked on the 6-digit 1993 *hhid*:

<i>hhid</i>	<i>hhid2004</i>	<i>pcode</i>
2220140	222014200	41
2220140	222014300	42
2380160	238016202	40
2380160	238016202	42

An individual can be reported by more than one of the households in a dynasty descending from an original 1993 household. This may be appropriate for household-based analyses, but for individual-based analyses, especially longitudinal ones, it is usually desirable to link each person with a single household in each wave in order to avoid double-counting.

For individuals who are only reported once this is straightforward. It also raises few difficulties if someone is only a member of only one household, or is a member of several households, but only resides in one of them. For individuals with multiple reports, but no memberships; multiple memberships but no residencies; or multiple residencies; deciding which household to assign them to involves an element of judgement.

The public release files include a longitudinally-linked file containing identifier variables. Individuals were allocated to households by successive application of a series of decision-making rules. Users should note that applying a different set of rules or applying the same rules in a different order would result in some individuals being allocated to different households. Data on resident members would be little affected and those on members as a whole affected less than those on individuals who are no longer a member of any household (who should be discarded from most analyses anyway).

The rules applied to eliminate multiple records for non-members or individuals who are members of multiple households are:

- Drop if resident in another household in same wave, but not the index one, or is a member of another household in the same wave, but not the index one.
- Drop 2004 members if the household cannot report the person's date of birth but the person belongs to another household that can.
- Drop multiple non-members in either 1998 or 2004 if the person is reported in the other wave, but only by other households, or if (s)he is only claimed as a member in the other wave by other household(s). Do the same for multiple members reported to be resident elsewhere in another wave.
- Drop non-members from households that do not include their mothers, fathers and spouses if other households do include them.
- Assign individuals who still have more than one record in 2004 to 'core' over 'next generation' and 'foster child' households.
- Drop individuals with more than one record in 2004 from the household that was interviewed first.
- Finally, assign anyone that still has two records in a wave to the '0' or '00' core household.

Linking individuals longitudinally reveals that some individuals reported as dead in 1998 were contacted in 2004 and that some individuals reported as alive in the 1998 wave were reported to have died before then in 2004. These inconsistencies should be resolved before working with the data longitudinally (we suggest accepting the 2004 report except when the person was actually resident or listed for the first time ($pcode \geq 40$) in 1998).

Of the people with data from both 1998 and 2004, 92% have ages that are consistent to within 3 or fewer years. 45 household members (0.86%) changed their reported gender between 1998 and 2004. As 42 of them have reported ages that are consistent to within 3 or fewer years, most are probably the same people.

Ian Timæus

Appendix: Edits made to the raw roster data

The following edits were made that reduce the number of records from that in the raw data file to that in the release version of the roster files:

- 1) All “null” records in the roster and section 1.3 file with a *pcode* of -2 were dropped. One such null record existed for each household which gained no new members in 1998-2004 (i.e. with a $pcode \geq 100$) and for each household where nobody died.
- 2) The records were dropped for 4923 prelisted individuals in *k* and *n* households who had not lived in those households in the previous year with the exception that 9 records were kept for decision makers mentioned in section 1.4.
- 3) One set of roster records was dropped for each of the 4 households that were interviewed twice (hhid2004: 223016303, 234032004, 236006005, 241004001).
- 4) The 274 records for non-members who died before 1998 were dropped.
- 5) Records on those non-members who are members of another household and are not mentioned in section 1.4 were also dropped.
- 6) All null and duplicated reports on deaths were dropped, usually keeping the report made by the dead person's 1998 household. Also, 7 people were reported as dead by one split household and alive by another: inspection of the questionnaires suggests that 4 of them were dead, but 3 of them alive. In all 7 cases, the misclassified records have been deleted. Similarly, 1 record for a person whose household “do not know” whether she is alive was deleted as another household reported her dead.

Group of records	Unedited roster records
<i>pcode</i> = -2 (not real people)	314
Non-members pre-listed in <i>k</i> and <i>n</i> HHs	4933
Dead before 1998 wave	274
Ex-members reported by core HHs – alive	3433
Ex-members reported by core HHs – survival not known	92
Deaths since 1998 wave	607
Living members from previous waves	5725
Living new members	3666
Total	19044

- 7) 19 duplicated records with *p*codes of 40-72 were dropped from the 2004 roster. These are records from dynasties that had split by 1998. They were dropped from the rosters of any split household to which the person did not belong in 1998.

- 8) A valid roster record and household cover record was created for each death report on someone from an extinct core household. *qtype* was set to 0 for these roster records.
- 9) Section 1.3 records with all variables set to missing (-4) were created for 20 deaths reported in 2004 as missed in 1998, but where the person was actually resident or was enumerated for the first time in 1998 and therefore almost certainly died later.

Other edits made to the raw data to produce Version 1.1 include:

- 1) *pcode*s were altered for mis-identified individuals in several households:

<i>hhid2004</i>	<i>pcode on questionnaire</i>		<i>pcode in version 1.1</i>
80003000	6	→	104
221011040	1	→	115
238016020	43	→	9
238016040	44	→	43
243006000	41	→	103

Corresponding changes to the *relhd93* and *mother* and *father* variables were made where necessary.

- 2) Some other variables were tidied up, a few data entry errors corrected, and some label conflicts were resolved.