

Appendix 1: Working with the relationship matrix

In order to simplify the use of the household matrix information a single compound string variable has been separated into 15 relationship variables rel1-rel15. Rel1 stores the respondent's relationship to person 1 and so forth.

Data availability

The Special Licence Household SAR contains information about the first 5 relationships recorded. This information is that which was collected on the main household form. A known bug on this variable means that rel5 is missing for households where the relationship matrix information was imputed.

The household CAM file contains all household matrix collected – however users need to be aware that the full household matrix information does not give full relationship information for larger households. The census household form only provides space for five individuals, larger households were required to use a continuation form for additional individuals. Individuals recorded on this continuation form did not record their relationship to all previous members of the household; rather they were asked only for their relationship to person 1 and to the 2 individuals who directly precede them. This means, for instance, that person number 6 recorded their relationship to persons 1, 4 and 5.

The Census offices' relationship algorithm has produced reciprocal relationships where appropriate. So for example, person 3 was asked to record their relationship to person 2. However, person 2 was not asked to record their relationship to person 3. Nevertheless, the Census offices have computed person 2's relationship to person 3 on the basis of the information given by person 3. So if person 3 recorded that they was the child of person 2, person 2 is computed to be the parent of person 3. As all household members record their relationship to person 1 this also means that person 1 has a full set of relationship to all household members.

Information held in the household variables is therefore given in the table below.

	Person number (persid)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rel1	self	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rel2	✓	self	✓	✓	✓	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
Rel3	✓	✓	self	✓	✓	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
Rel4	✓	✓	✓	self	✓	✓	-9	-9	-9	-9	-9	-9	-9	-9	-9
Rel5	✓	✓	✓	✓	self	✓	✓	-9	-9	-9	-9	-9	-9	-9	-9
Rel6	✓	-9	-9	✓	✓	self	✓	✓	-9	-9	-9	-9	-9	-9	-9
Rel7	✓	-9	-9	-9	✓	✓	self	✓	✓	-9	-9	-9	-9	-9	-9
Rel8	✓	-9	-9	-9	-9	✓	✓	self	✓	✓	-9	-9	-9	-9	-9
Rel9	✓	-9	-9	-9	-9	-9	✓	✓	self	✓	✓	-9	-9	-9	-9
Rel10	✓	-9	-9	-9	-9	-9	-9	✓	✓	self	✓	✓	-9	-9	-9
Rel11	✓	-9	-9	-9	-9	-9	-9	-9	✓	✓	self	✓	✓	-9	-9
Rel12	✓	-9	-9	-9	-9	-9	-9	-9	-9	✓	✓	self	✓	✓	-9
Rel13	✓	-9	-9	-9	-9	-9	-9	-9	-9	-9	✓	✓	self	✓	✓
Rel14	✓	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	✓	✓	self	✓
Rel15	✓	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	✓	✓	self

- ✓ Indicates that the data is available either because it was collected directly or was calculated reciprocally
- 9 Indicates that the data was not collected
- Self is recorded in the data and is coded as -8
- The shaded are indicates information collected by the main household form. This information, if appropriate is available in the SL-HSAR.
- The white are indicates data collected using continuation form(s). This information, is only available on the CAM version of the data.

The maximum number of individuals a household is 15. Most households are much smaller and therefore a large number of relationships are not applicable. A household of one will not have applicable values for rel2-rel15 (and rel1 will be recorded as 'self'). Where a relationship is inapplicable because no such relationship exists (due to household size) these relationship variables will be coded as -7.

Interpretation of the variables

In order to interpret the variable consider this as <person n> is the <rel*> of <*>. An example: Person 2 has a value of rel1 indicating "child". This means that person 2 is the child of person 1.

Data quality

We know that the relationship matrix was considered especially problematic by respondents. Although the edit and imputation procedure was able to detect anomalous responses which suggested that relationships had been reversed (e.g. where a parent recorded their relationship to their child as 'child' not 'parent'), a small number of relationships could not be dealt with by the census offices' relationship algorithm. Furthermore, additional anomalous relationships might have resulted from the imputation of additional members of the household as part of the One Number Census procedure. The data is not as clean, therefore, as census users might normally expect (Office for National Statistics 2003a).

Data validation procedures suggest that one outcome of imperfect relationship data is an inflation in the number of same sex couples in the data. The following information is quoted from the Census Data validation report (ONS 2003b).

Same Sex Couples

The validation process detected that there was an inflated number of same sex couples, both with and without children. The main reasons were:

- the form filler ticked the wrong sex for one person;
- duplicate instances of the same person;
- Partner' being imputed when the relationship matrix was incomplete or inconsistent;
- the One Number Census process adding a person into an existing household; and
- the Household Composition Algorithm deriving an incorrect family type.

A set of automated checks, including analysis of first names to check whether the wrong sex had been ticked, was carried out to identify people who fell into the above categories. Where couples could not be decided automatically the record's validity was decided manually. A few households containing more than one potential same sex couple were passed straight to manual checking.

The number of same sex couples counted in the Census in England and Wales was 88,322. Of these, 39,261 were found to be genuine (44.5 per cent). A further 18.0 per cent were imputed by the ONC process, 12.3 per cent were caused by the wrong sex being ticked, and 25.2 per cent were due to imputed relationship information. This information was collected to produce a Same Sex Couples database. To date, no data has been corrected or flagged on the Census Output Database.

References:

Office for National Statistics (2003a) **Census 2001 Review and Evaluation Edit and Imputation Evaluation Report**

<http://www.statistics.gov.uk/census2001/pdfs/editimputevr.pdf> last accessed 02/08/06

Office for National Statistics (2003b) **Census 2001 Review and Evaluation Data Validation: Executive Summary**

http://www.statistics.gov.uk/census2001/pdfs/data_validation_es.pdf last accessed 02/08/06