National Records of Scotland

2011 Census

Release 2C - Household Relationships Data Quality Issues

December 2013

Household Relationships Data Quality Issues

The relationship question was the most complex question in the census; it was a matrix style question that collected the relationships of persons recorded on the questionnaire to each other. For large households which required a continuation questionnaire the additional persons were linked to person one on the main household questionnaire, as well as to the previous two people in the household.

A selection of deterministic edits were applied to the relationship data prior to imputation, where the correct response could be deduced with a high degree of certainty. Following this, and the subsequent probabilistic imputation, the relationship data was complete, and largely consistent. However we were aware that certain implausible/impossible relationships still remained.

The quality of the relationship information was reduced following adjustment for under enumeration. The Census Coverage Adjustment process added people into existing households to adjust for person non-response. Relationships for these new people were created using the same process as had been applied to the enumerated records, but in a small number of cases this resulted in improbable household relationships being created. For example, the Adjustment process might add a married adult into a household containing an existing married couple. Because the edit rules specify that one person cannot have two spouses, or a spouse and a partner, the imputation process might create a partnership between the additional person and an existing child in the household, or between the additional person, and a single elderly relative living in the house.

Following this adjustment, the relationship information was used to populate 6 output variables which together describe the nature of the household composition (Household Composition Algorithm). However, because some unusual relationships still existed in the data, the outputs of the HCA contained inconsistencies and for this reason the relationships and household composition variables in 3500 households containing one or more of the following issues were manually amended:

1) people assigned to the wrong generation
2) family or household reference persons who were aged under 16
3) couples where one or both members of the couple was aged under 16
4) people aged 80+ who were the child of another household member
5) parents less than 14 years older than their children
6) mothers more than 49 years older than their children
7) 22 households with impossible relationships which remained after the initial run through Edit & Imputation.

These edits resolved the majority of the inconsistencies in the data which we were aware of. In a small number of cases where it looked as though the unusual relationships were genuine, and where allowing the data to remain unaltered would be unlikely to produce issues with future outputs, we have not made any changes.

More information on the reasons for the inconsistencies found in the data will be reported in a subsequent report as part of the evaluation.

The following release 2C variables may be affected by the household relationships data quality issues:

ADEMHUK11
ADULTLSPUK11
AHTHUK11
DCHPUK11
DPCFAMUK11
DPCHUK11
ELARPUK11
FMSPS11
HHCHUK11
HHLSHUK11
HRPPUK11
LARPUK11
MEIGHS11
NSSHUK11
OCCUPANCY RATING