

TTP DATA QUALITY REPORT

DATA SOURCE

DATA SUPPLIER:	Census Office
DESCRIPTION:	2001 Census (Entire enumerated population where subject is a student at term time address or subject is not a student) taken on 29 th April 2001.
COVERAGE:	Northern Ireland
NUMBER OF RECORDS:	1,603,641 ¹ http://www.nisra.gov.uk/Census/2001Census.html 1. This varies from the full Census count as it does not include imputed/CUE records

DATA BACKGROUND

A census of population and housing, even in a country as small as Northern Ireland, is a major undertaking and much time, effort and resources are devoted to it. The Census in Northern Ireland is taken under the authority of the Census Act (Northern Ireland) 1969, as amended by the Census (Confidentiality)(Northern Ireland) Order 1991, which gives powers for taking a census from time to time provided that, at the commencement of the year in which a census is to be taken, at least five years have elapsed since the commencement of the year in which a census was last taken. The Act also gives powers for the census to be taken on a particular day.

The 2001 Census was taken on 29 April 2001 and placed a legal obligation on every household in which someone was usually resident on Census Day, and on every person who was a usual resident of a communal establishment, to complete a Census form. This contrasts to the 1991 Census which collected information on both usual residents and visitors present on Census night.

Students and school children in full-time education studying away from home were enumerated as resident at their term-time address. Basic demographic information only (name, sex, age, marital status and relationship) was collected at their 'home' or vacation address. HM Forces permanently posted in Northern Ireland were enumerated at their actual address of residence unless they were married and unaccompanied by their spouse. In these cases they were included at the address they shared with their spouse. In contrast, HM Forces posted in Northern Ireland on short term tours of 3 – 4 months or less (Roulement troops) were not included in the Northern Ireland Census. Arrangements were made for these troops to be enumerated at their home garrisons.

It was not possible to enumerate all persons by the conventional method, and special arrangements had to be made for certain population groups, such as the Armed Forces, prisoners, students, shipping personnel, refugees and asylum seekers, the visually impaired, persons sleeping rough, travelling fairs and circuses, and others, though standard forms were used in all cases.

Also, a series of new procedures, based around the basic enumeration procedure, was required at short notice to deal with the outbreak of Foot and Mouth disease that had spread to large parts of the country before the enumeration was due to take place. Special methods of delivery and collection were quickly arranged to ensure that as complete coverage as possible was achieved in the areas hit by this major epidemic.

The Census was followed by a Census Coverage Survey (CCS), an independent count of people in a sample of small areas throughout Northern Ireland. The results of the CCS were combined with the data collected from the Census in order to allow estimates of households and people missed by the Census to be made. The figures have been adjusted to allow for under-enumeration.

The 2001 Census processing system used Optical Mark Recognition (OMR) and Optical Character Recognition (OCR) to lift the data from the completed Census forms and store it electronically.

It should be noted that the number of records quoted relates to the entire enumerated population. It does not include those individuals who have been imputed which accounts for 5% of the total records in Census 2001. If matching to a 100% complete administrative dataset the maximum match rate that can be achieved is 95%.

DEMOGRAPHIC INFORMATION

Information available to the TTP for linkage purposes -

- Name information which includes Forename, Midname and Surname
- Address information at the time of the Census
- Sex
- Date of birth
- Pointer UPRN

Name, Address, Sex, Date of birth and UPRN are also available for other household members. This can be used for matching purposes and linkage quality checks.

DATA QUALITY & PRE-PROCESSING

Record linkage is highly dependent on the quality of the data being linked. Many key variables can be presented quite differently between datasets as the information has been collected and recorded using different methods. This can greatly complicate record linkage unless understood ahead of time. Pre-processing can be used to standardise these variables into a consistent format prior to linkage.

It should be noted that name and date of birth information can be more reliable when extracted from administrative data sources where regular checks are completed such as Health Card Registrations. This is due to the Census form being self-completed and scanning errors caused by unrecognisable characters and poor handwriting.

NAME INFORMATION- Name information is available for 99.9% of records. Midname is populated where applicable.

- Punctuation and characters which do not appear in the English alphabet (e.g. Á, %, \$...) can appear in the name information. A function can be used to replace these with usable characters;

ADDRESS INFORMATION- House number, street name and town is available for approximately 97.8% of the total records. There are a small number of records (<0.01%) that do not have postcode information.

- Postcode information is standardised prior to linkage.

SEX- Information relating to the sex of the individual is available for 99.6% of records.

- Sex information should be standardised prior to linkage.

DATE OF BIRTH- Information is available for 99.6% of records.

- Date of birth can be parsed into its constituent parts (day, month and year) prior to linkage

POINTER UPRN- Information is available for 96% of records.

- Pre-processing is not required for this data.

MATCHING METHODS

Using the demographic information provided the following matching can be undertaken-

- Person matching
- Address matching
- Associative matching of other household members

PREVIOUS LINKAGE MATCH RATES

The Census 2001 information has previously been matched to health data as part of the Northern Ireland Longitudinal Study (NILS). This typically produces an adjusted match rate of approximately 99.7%; More information can be found on the [NILS Metadata](#).

EVALUATION

The Census 2001 information is generally of good quality and can be used for record linkage easily and without extensive pre-processing.

It should be noted that when matching Census 2001 to a 100% Northern Ireland dataset we will be unable to match a small number of records which relate to individuals who were imputed in the Census data and for which we hold no demographic information (approximately 5%).