

IMPUTATION FLAG VARIABLES

Introduction

The census endeavours to be the most complete source of information about the England and Wales population. However, similarly to other self-completion questionnaires, respondents sometimes make errors when recording their answers, resulting in data that cannot be used for estimation and analysis. Invalid responses include blanks, multi-ticks, out of range values and partially answered questions. A respondent's answers may also be inconsistent, e.g. if a respondent gave their age as five years old and an occupation, the two values would be considered inconsistent because by definition a five-year cannot work. Invalid responses may be intentional (e.g. the respondent does not want to answer the question or does not know the answer) or unintentional (e.g. the respondent thinks that they can tick more than one option or misses a question).

ONS has developed several methods that have been applied to resolve item non-response, invalid responses, and item inconsistencies in the 2001 and 2011 Census data. These methods included deterministic editing and imputation. Deterministic editing resolves inconsistent or invalid responses by determining the correct value from other valid values on record. Three questions in the 2011 Census were edited this way, marital and civil partnership status, position in establishment, and relationship to other household members.

Imputation is a statistical process that replaces invalid or missing responses to a question with valid values. Two types of imputation were used for the census:

- nearest neighbour donor imputation: this was the commonest type used. This method matches respondents with invalid response values ("recipients") to those with valid response values ("donors") based on various characteristics, and the valid response values from the "donor" with the closest overall match to the "recipient" in terms of characteristics, are given to the "recipient".
- manual imputation: this was only used in a small number of cases, where nearest neighbour donor imputation did not fully resolve invalid or missing responses.

More information on the census item and imputation process is available from the [ONS website](#).

This document aims to introduce researchers to the imputation indicator variables that are available for the 2001 and 2011 census files of the ONS Longitudinal Study (LS). The variables are provided for both LS members and non-members¹.

What are the imputation indicator variables?

The 2001 and 2011 Census files for member (ME01 and ME11) and non-members (NM01 and NM11) contain a series of variables that indicate if a variable has been edited or imputed, the imputation

¹ Individuals living in the same household as an LS member at the time of the census.

indicator variables. In ME01 and NM01 the imputation flag variables all end with “IMP”, and in ME11 and NM11 they end with “_IMP”.

For the 2001 Census files in the LS, the imputation indicator variables were binary:

0 = Not Imputed

1 = Imputed

For 2011 they had four categories:

0 = Not Imputed

1 = Edited/Imputed from a missing value

2 = Edited/Imputed from a valid value (including multi-ticks)

-9 = No retained record

We recommend that researchers use the imputation indicator variables to see if the information that their outcome or main exposure indicators are based on is imputed. Based on this, researchers can decide whether to exclude those LS members whose data has been imputed. If the decision is taken to include LS members with imputed outcome or main exposure indicators, we also recommend that the analyses is repeated by excluding these individuals.

Tables 1 and 2 below list the imputation indicator variables that are included in the 2001 and 2011 Census files. In 2001, excluding the 30 variables indicating whether the LS member’s relationship to person 1 – 30 in the household is imputed (REL01IMP – REL30IMP), there are 78 imputation indicator variables, and in 2011 there are 46.

*Table 1: imputation flag variables for the 2001 Census for LS members**

Variable name	Description
ACCHIMP	Type of accommodation imputation indicator. 2001.
ACTLWIMP	Activity last week imputation indicator. 2001.
ADEMIMP	Adults in employment in household imputation indicator. 2001.
ADTHIMP	Adults in household imputation indicator. 2001.
AFRHIMP	Household reference person in the armed forces imputation indicator. 2001.
AGEPIMP	Age imputation indicator. 2001.
AHCHIMP	Alternative household composition imputation indicator. 2001.
AHTHIMP	Alternative household type imputation indicator. 2001.
BSTHIMP	Availability of bath/shower imputation indicator. 2001.
CAVHIMP	Number of cars imputation indicator. 2001.
CEAHIMP	Carers in household by economic activity imputation indicator. 2001.
CNHHIMP	Central heating imputation indicator. 2001.
COBHIMP	Household reference person - country of birth imputation indicator. 2001.
COBPIMP	Country of birth imputation indicator. 2001.
CRSHIMP	Number of carers in household imputation indicator. 2001.
DCHPIMP	Dependent child imputation indicator. 2001.
DEPEDIMP	Education deprivation imputation indicator. 2001.
DEPEMIMP	Employment deprivation imputation indicator. 2001.
DEPHDIMP	Health and disability deprivation imputation indicator. 2001.

Variable name	Description
DEPHSIMP	Housing deprivation imputation indicator. 2001.
DOBIMP	Date of birth imputation indicator. 2001.
DPCFMIMP	Number of dependent children in family imputation indicator. 2001.
DPCHIMP	Number of dependent children in household imputation indicator. 2001.
DTWPIMP	Distance travelled to work imputation indicator. 2001.
ECOIMP	Economic activity imputation indicator. 2001.
EMTPIMP	Employment status imputation indicator. 2001.
ENPCIMP	Enumeration postcode imputation indicator. 2001.
ETHHIMP	Household reference person - ethnic code imputation indicator. 2001.
ETHPIMP	Ethnicity imputation indicator. 2001.
EVERWIMP	Ever worked imputation indicator. 2001.
FMSPIMP	Family status imputation indicator. 2001.
FRPPIMP	Family reference person imputation indicator. 2001.
HEAPIMP	Health imputation indicator. 2001.
HELPIMP	Carer imputation indicator. 2001.
HHCHIMP	Household composition imputation indicator. 2001.
HLQPIMP	Highest qualification imputation indicator. 2001.
HOUPIMP	Hours worked imputation indicator. 2001.
HRPPIMP	Household reference person imputation indicator. 2001.
ILAHIMP	Limiting long term illness in the household by age imputation indicator. 2001.
ILLHIMP	Limiting long term illness in household imputation indicator. 2001.
ILLPIMP	Limiting long term illness imputation indicator. 2001.
INDPIMP	Industry imputation indicator. 2001.
LANHIMP	Landlord imputation indicator. 2001.
LANIMP	Welsh language imputation indicator. 2001.
LARPIMP	Living arrangements imputation indicator. 2001.
LFLHIMP	Lowest floor level of accommodation imputation indicator. 2001.
LWYPIMP	Year last worked imputation indicator. 2001.
MEIGHIMP	Multiple ethnic identities imputation indicator. 2001.
METHIMP	Multiple ethnicity in England & Wales imputation indicator. 2001.
MGRPIMP	Moving group reference person imputation indicator. 2001.
MIGHIMP	Wholly moving household imputation indicator. 2001.
MIGPIMP	Usual address 1 year ago imputation indicator. 2001.
MSTPIMP	Marital status imputation indicator. 2001.
MTWPIMP	Mode of transport to work imputation indicator. 2001.
NORHIMP	Number of rooms in household imputation indicator. 2001.
NSSHIMP	Household reference person NSSEC imputation indicator. 2001.
NSSPIMP	National socio-economic classification (NSSEC) imputation indicator. 2001.
OCCPIMP	Occupation imputation indicator. 2001.
OWNRNIMP	Owns/rents imputation indicator. 2001.
PENHIMP	Pensioner household imputation indicator. 2001.
PENPIMP	Pensionable age imputation indicator. 2001.
POSPIMP	Position in establishment imputation indicator. 2001.
POWPIMP	Place of work imputation indicator. 2001.
PQUPIMP	Professional qualifications imputation indicator. 2001.
QUPIMP	Qualifications imputation indicator. 2001.
RELPIMP	Person relationship imputation indicator. 2001.
REL01IMP – REL30IMP	Relationship to person 1-30 imputation indicator01
RMRHIMP	Rooms required imputation indicator. 2001.

Variable name	Description
SCAHIMP	Self-contained imputation indicator. 2001.
SEXIMP	Sex imputation indicator. 2001.
SIZPIMP	Company size imputation indicator. 2001.
STAHIMP	Students living away imputation indicator. 2001.
STAPIMP	Student accommodation imputation indicator. 2001.
STUPIMP	Student imputation indicator. 2001.
SUMPIMP	Supervisor/manager imputation indicator. 2001.
TENHIMP	Household tenure imputation indicator. 2001.
TPNIMP0	Total people needing impute. 2001.
TTIMP	Term time address imputation indicator. 2001.
UA1YAIMP	Usual address 1 year ago imputation indicator. 2001.
WG1FMIMP	Workers in generation 1 of family imputation indicator. 2001.
WHCHIMP	Welsh household composition imputation indicator. 2001.
WTRHIMP	Workers in household and their transport imputation indicator. 2001

*Equivalent imputation flag indicators exist for non-members (i.e. individuals in the same household as the LS member on census night) in the non-member file for 2001 (NM01). The equivalent variables have the same variable name prefixed with "N" e.g. NAGEPIMP = age of non-member imputation indicator, 2001.

Table 2: imputation flag variables for the 2011 Census for LS members *

Variable name	Description
ACCH11_IMP	Type of accommodation imputation indicator. 2011.
ACTLW11_IMP	Activity last week imputation indicator. 2011.
AGEP11_IMP	Age imputation indicator. 2011.
ARRXMNTHAGO11_IMP	Number of months since arrival to the UK imputation indicator. 2011.
CAVH11_IMP	Number of cars imputation indicator. 2011.
CNHH11_IMP	Type of central heating imputation indicator. 2011.
COBP11_IMP	Country of birth imputation indicator. 2011.
EMTP11_IMP	Employment type imputation indicator. 2011.
ETHP11_IMP	Ethnic group imputation indicator. 2011.
EVERW11_IMP	Ever worked imputation flag. 2011.
HEAP11_IMP	General health imputation indicator. 2011.
HELP11_IMP	Provision of unpaid care imputation indicator. 2011.
HOUP11_IMP	Hours worked imputation indicator. 2011.
IDENINT11_IMP	National identity (text) imputation indicator. 2011.
IDENUK11_IMP	National identity (tick) imputation indicator. 2011.
ILLP11_IMP	Long-term health problem or disability imputation indicator. 2011.
INDP11_IMP	Industry Imputation Indicator. 2011.
INTENTION11_IMP	Intention to stay imputation indicator. 2011.
LANGPRF11_IMP	English proficiency imputation indicator. 2011.
LANH11_IMP	Landlord imputation indicator. 2011.
LANP11_IMP	Welsh language imputation indicator 2011.
LWYP11_IMP	Year last worked imputation indicator. 2011.
MAINLANG11_IMP	Main language imputation indicator. 2011.
MSTP11_IMP	Marital status imputation indicator. 2011.
MTWP11_IMP	Mode of transport for work imputation indicator. 2011.
NOBH11_IMP	Number of bedrooms imputation indicator. 2011.
NORH11_IMP	Number of rooms in the household imputation indicator. 2011.
OCCP11_IMP	Occupation imputation indicator. 2011.
OWNREN11_IMP	Owens or rents imputation indicator. 2011.
POSP11_IMP	Position in Communal Establishment imputation indicator. 2011.

Variable name	Description
POWPEW11_IMP	Place of work indicator imputation indicator. 2011.
PSPTTEL11_IMP	Passports held imputation indicator. 2011.
PSSPRT11_IMP	Imputation indicator for PASSPUK111, PASSPUK112, PASSPUK113 and PASSPUK114.
QUP11_IMP	Qualifications imputation indicator. 2011.
REL01IMP11 – REL30IMP11	Relationship to person 1 – 30 imputation indicator. 2011.
SCADDTYP11_IMP	Second address type imputation indicator. 2011.
SCADIND11_IMP	Second address indicator imputation indicator. 2011.
SCADINT11_IMP	Second address Country imputation indicator. 2011.
SCAH11_IMP	Self-contained accommodation imputation indicator. 2011.
SEX11_IMP	Sex imputation indicator. 2011.
STUP11_IMP	Student indicator imputation indicator. 2011.
TTIND11_IMP	Term-time address indicator imputation indicator. 2011.
WKPLINT11_IMP	International workplace address indicator imputation indicator. 2011.
WPCIND11_IMP	Workplace indicator imputation indicator. 2011.
YRADIND11_IMP	Address one year ago Response indicator imputation indicator. 2011
YRADINT11_IMP	Country of residence one year ago (non-UK) imputation indicator. 2011.
YRARRMON11_IMP	Month of arrival in the UK imputation indicator. 2011.
YRARYEAR11_IMP	Year of arrival in the UK imputation indicator. 2011.

*Equivalent imputation flag indicators exist for non-members (i.e. individuals in the same household as the LS member on census night) in the Non-member file for 2011 (NM11). The equivalent variables have the same variable name prefixed with "N" e.g. NAGEP11_IMP = age of non-member imputation indicator. 2011.

Number of LS members with imputed information

Approximately 57% of LS members do not have any imputed information for the 2001 Census, 25% have between 1 and 10 imputed variables and approximately 11% have 11-20 imputed variables (see tables 1a and 1b in Sheet 1 of the associated Imputed variables spreadsheet). In 2011, approximately 66% of LS members do not have any imputed information and approximately 32% have between 1 and 10 imputed variables (see tables 2a and 2b in Sheet 1 of the associated Imputed variables spreadsheet).

Figure 1 below shows the 20 most frequently imputed variables from the 2001 Census (table 3 in Sheet 2 of the associated Imputed variables spreadsheet shows the same information but for all of the imputation indicator variables). The most frequently imputed variable was the household reference person indicator (17.8%), followed by the housing and employment deprivation indicators (16.5% and 15.6% respectively). The sex, Welsh language, marital status, age and date of birth variables were imputed for the least number of individuals (<0.5%).

Figure 1: The 20 most frequently imputed variables from the 2001 Census

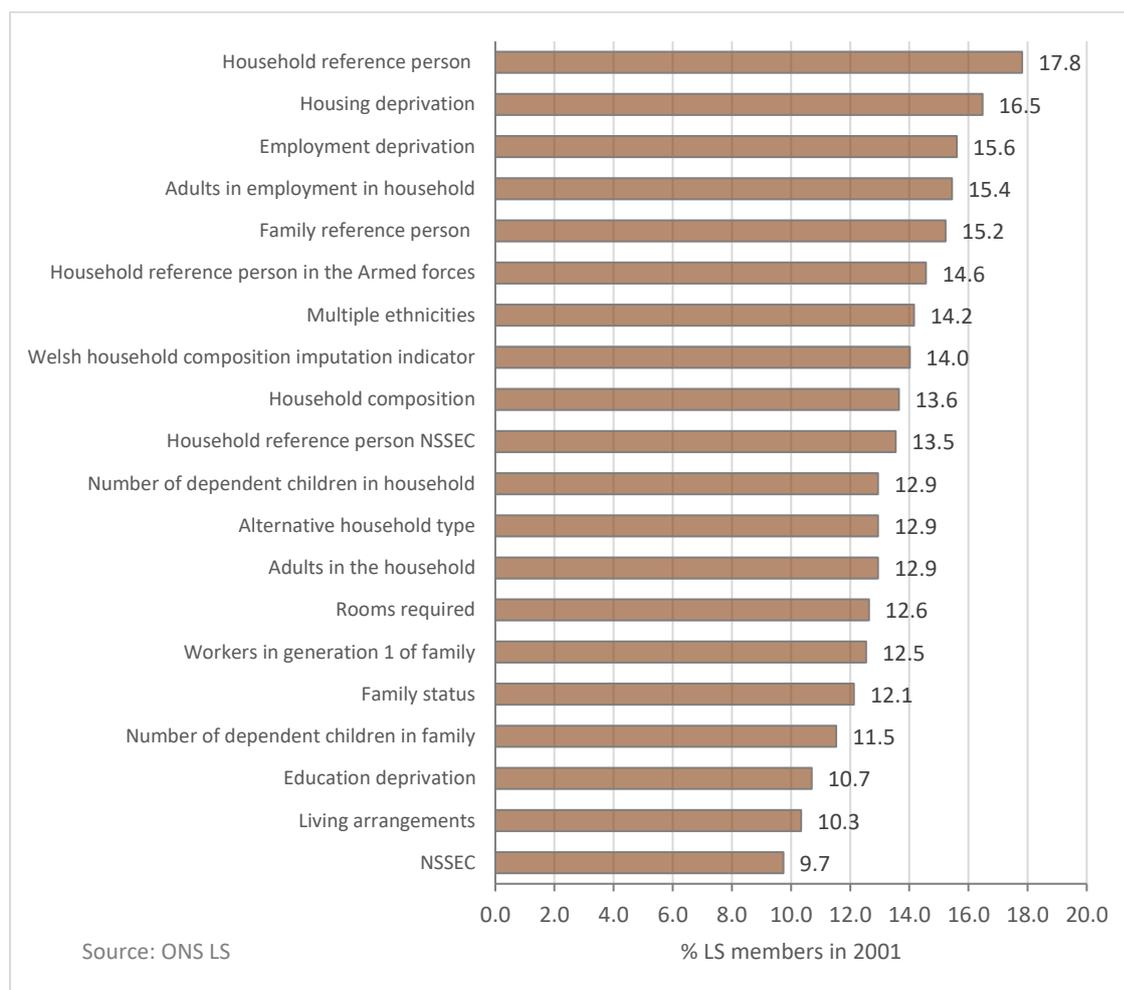


Figure 2 below (and Table 4 in Sheet 3 of the associated Imputed variables spreadsheet) shows the same information for the 2011 variables. The most frequently imputed variable in 2011 was the industry variable (8.8%), followed by the workplace indicator, international workplace indicator and place of work variables (5.8%, 5.6% and 5.3% respectively). The sex, position in communal establishment, age and Welsh language variables were imputed for the least number of individuals (<0.5%).

Figure 2: The 20 most frequently imputed variables from the 2011 Census

