



# Item Non-response, Item Deterministic Editing and Item Imputation

December 2012

This is one of a series of reports published to support the release of results from the 2011 Census. This series of methods and quality reports provides information on the different methods used to collect, process, clean, adjust and protect the census results. The series also reports on the quality assurance of the results and provides quality indicators.

Terms used in the series are explained in the [2011 Census glossary](#).

## Contents

---

Summary .....	2
Background and introduction .....	2
Item non-response, item deterministic editing and item imputation .....	3
Population base and denominators .....	5
Comparison with 2001 .....	5
Comparison with Northern Ireland and Scotland .....	6
Comparison with other sources .....	7
Annex A: Variables included in item rates.....	8
Annex B: Eligibility for inclusion in item rate denominators.....	10

# 1. Summary

This paper outlines the definitions of the item non-response rate, item editing rate and item imputation rate for the 2011 Census of England and Wales.

These rates are provided down to local authority level and are [available to download](#) from the census website.

Further information on the [item edit and imputation process](#) is also available to download.

# 2. Background and introduction

In 2011 the England and Wales census questionnaire was sent out to every home and delivered to every communal establishment in England and Wales. The residents could either complete the paper questionnaire provided, or complete online. As with any self-completion questionnaire, it was common for the respondents to make errors when recording their answers, resulting in data that were not valid for estimation and analysis. Referred to as item non-response, this can be unintentional, for example where a respondent misses a question or thinks they can tick more than one option, or intentional where a respondent either does not know the answer or does not want to provide the answer.

It was also common for some correctly recorded values to be considered invalid because they were inconsistent either with other values on the questionnaire, or with auxiliary information or definitions. Referred to as item inconsistency, these errors were detected by validating the data against a set of predefined edit rules. For example, the rule which stated that a person aged less than 16 years old could not have a qualification would have flagged a record where a person gave their age as five and said they had a university degree. The 2011 Census edit rules are available in the item edit and imputation process paper.

The 2011 Census data was edited in two stages to correct for item non-response and inconsistencies. Firstly, a limited number of deterministic edits were applied. This was followed by item imputation which corrected for non-response and inconsistencies. The purpose of this paper is to define how the rates of item non-response, item deterministic editing and item imputation have been calculated. The definition for each rate is given in section three; section four defines the population base and denominators and sections five, six and seven give a brief statement on comparability to the 2001 Census, rest of UK and other sources respectively.

## 3. Item non-response, item deterministic editing and item imputation

This section outlines the broad definitions for each rate calculated for the 2011 Census. The population base and denominators are defined in section four.

### **Item non-response rate**

Item non-response refers to an invalid response to a question in a completed questionnaire. Invalid responses include blanks, multi-ticks, out of range values and partially answered responses, for example in second address, occupation or industry which are collected in more than one field.

The item non-response rate is the percentage of the measured population that had an invalid value for that item. It is calculated by dividing the total number of invalid responses for an item by the total number of persons who were required to answer that item.

### **Item deterministic editing rate**

Deterministic editing is a process used to resolve invalid or inconsistent values by determining the correct value directly from other values in the record. Only three questions were edited deterministically in the 2011 Census; marital and civil partnership status, position in establishment and relationship. Deterministic editing was used prior to item imputation and it should be noted that values determined by this process may have been changed again by imputation in order to resolve inconsistencies with other variables.

The item deterministic editing rate is the percentage of the measured population that had a value edited deterministically for that item. It is calculated by dividing the total number of values deterministically edited for an item by the total number of persons who were required to answer the question.

### **Item imputation rate**

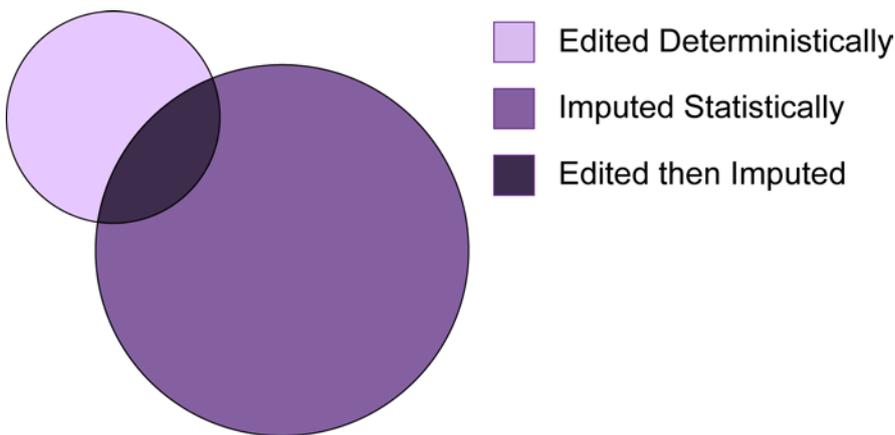
Item imputation is a statistical process that replaces invalid responses to a question with a valid value. It can also be used to replace valid values that are found to be inconsistent according to predetermined definitions or edit rules.

The item imputation rate is the percentage of the measured population whose values have been changed by the imputation process. It is calculated by dividing the total number of imputed responses by the total number of persons who were required to answer that item.

The difference between the item non-response rate and the item imputation rate is the item inconsistency rate. This is the percentage of responses that were replaced due to failing the edit rules, such that:

**Item imputation rate = Item non-response rate + Item inconsistency rate**

The item inconsistency rate is not reported separately in tables, but can be derived from the difference between the item imputation rate and the item non-response rate. As shown in figure 1, some values were edited by item deterministic editing and then changed again by item imputation. This occurred if the edited value was inconsistent with another value in the record, or with a value subsequently imputed into the record. Therefore the sum of the item deterministic editing rate and the item imputation rate slightly overstates the total imputation for the item.



**Figure 1: Overlap in the item deterministic editing and imputation rates**

## 4. Population base and denominators

The population base for item editing and imputation is all responding individuals, including short term residents. It is smaller than the final census population base because it does not include persons added by the [Census Coverage Adjustment Process](#). Students who live elsewhere during term time are included in the items that they answered on the questionnaire; age, sex, marital and civil partnership status, second address, type of second address and student.

The denominator for each item depends on the routing on the questionnaire. For example, persons aged less than 16 years old did not answer the qualifications and employment questions and are excluded from the denominator of those items. A list of the standard variables reported is provided in Annex A and the rules for who is included in each denominator are given in Annex B. For each item the denominator is the same across all three rates. Eligibility is assessed on the final values after editing and imputation.

## 5. Comparison with 2001

While the rates for the England and Wales 2011 Census have been calculated using the same principles that were applied in the 2001 Census there are several methodological differences that mean the rates are not directly comparable. Information for 2001 is available in the [2001 Census Edit and Imputation evaluation report](#).

### Differences in editing

The main difference lies in the application of deterministic edits prior to imputation. In 2001, 13.7 million edits were applied to 11.8 million people. The most frequently applied edits were to amend qualifications, provision of unpaid care, method of travel to work, supervisor status, general health and country of birth. In addition the marital status of those under the age of 16 was set to single (never married) if missing.

In 2011 deterministic editing was restricted to three questions which tended to have a high level of response error; marital and civil partnership status, position in communal establishment and household relationships.

For marital and civil partnership status, an edit similar to the 2001 edit was applied that set missing marital and civil partnership status to 'never married or in a civil partnership' for persons who did not record a spouse, civil partner, or have any missing relationships in the household relationships table. As in 2001, the majority of missing marital and civil partnership statuses were for those aged less than 16 years old.

Any communal persons identifying as 'staff' without 'working' in activity last week (including if it was missing) were changed from 'staff' to 'resident'.

The remaining edits applied were to the household relationships table. An algorithm was written to correct for the most common response errors and determine missing relationships where possible by triangulating observed values. For example, where parents ticked 'parent' for their child instead of 'son or daughter' the value was changed to 'son or daughter' and missing values between two persons who shared the same parents were changed to 'brother or sister. Further information on the relationship algorithm can be found in the item edit and imputation process paper.

### **Differences in denominators and population base**

The denominators for each rate include all individuals who were observed or imputed to require a response to the question. This was the same approach used in 2001 and for most questions the persons in scope is the same. However in 2001 the labour market questions were not answered by those aged 75 years or older and so the denominator in 2011 includes more individuals because all persons over 15 answered the labour market questions in 2011. This includes; qualifications, activity last week, ever worked, employment status, supervisor status, occupation, industry, workplace address, hours and method of travel to work.

## 6. Comparison with Northern Ireland and Scotland

The rates are harmonised where possible across England, Wales and Northern Ireland (NI). The rates for Scotland have not yet been produced and deviations for Scotland may be reported in the future.

Where a question was not asked in all countries, for example 'Welsh language', it will be reported as not applicable in joint tables. Where questions were asked in all countries they are reported using the same naming convention. Questions not asked in England and Wales include the equivalent Northern Ireland and Scottish language questions, health conditions (NI/Scotland) and religion brought up in (NI).

### **Differences in denominators and population base**

In Northern Ireland year and month of arrival are to Northern Ireland, not UK and are asked to all persons who have lived outside of Northern Ireland for one year or more. In England, Wales and Scotland the question is asked only to those born outside the UK. Therefore the denominators are different between the countries for the arrival in the UK question.

In Scotland and Northern Ireland the workplace address and travel to work questions were asked of all persons except students who live somewhere else during term time. In England and Wales these questions were only asked of those over the age of 15 and working the week before census night. So the denominator is smaller in England and Wales.

## 7. Comparison with other sources

The principles used to calculate these rates are similar to those applied to social surveys such as FRS (Family Resource Survey) and EUSILC (European Union Statistics on Income and Living Conditions) in that they give a breakdown of the final data. In some social surveys the rates are given at the aggregate level for the whole data, rather than by item, and the FRS for example only reports imputation where more than 30 values were amended. As social surveys are largely collected with computer assisted interviewing, edit checks occur at the time of response and imputation does not usually include consistency checks. Therefore the imputation rate will be the same or less than the nonresponse rate. The extent to which concepts and questions are the same varies from question to question and each needs to be considered individually.

The rates reported here are the same as those reported to EUROSTAT for population censuses.

Further information on how the 2011 Census compares to other data sources is [available to download](#).

## Annex A: Variables included in item rates

Variable name	Variable description
Actlw	Activity last week
Age	Age
Bedrooms	Number of bedrooms
Carer	Provision of unpaid care
Carsno	Number of cars or vans in household
Cenheat	Central heating
COB	Country of birth
Disability	Long-term health problem or disability
Empstat	Employment status
Ethnicity	Ethnic group
Everwork	Ever worked
Health	General health
Hours	Hours worked
Industry_Everworked	Industry where previously worked
Industry_Working	Industry where currently working
Intention	Intention to stay
Landlord	Landlord
Langprf	Proficiency in English
Lastyrwork	Last year worked
Mainlang	Main language
Marstat	Marital and civil partner status
National_Identity	National identity
Occ_Everworked	Occupation where previously worked
Occ_Working	Occupation where currently working
PersonCount	Person count
Position	Position in communal establishment
Psptel	Passports held
Pssprt	Passports held (UK)
Quals	Qualifications
Re1puk	Relationship to person one
Religion	Religion
Rooms	Number of rooms
Scaddtyp	Type of second address
Scadind	Second address (indicator)
Scadint	Second address (international)

Scadpc	Second address (postcode)
Selfcon	Self-contained
Sex	Sex
Student	Schoolchild or full-time student indicator
Supervisor	Supervisor status
Tenure	Tenure of household
Termind	Term-time address indicator
Transport	Method of travel to work
Typacom	Accommodation type
Welshlang	Welsh language
Wkplind	Workplace address (indicator)
Wkplint	Workplace address (international)
Wkplpc	Workplace address (postcode)
Yradind	Address one year ago (indicator)
Yradint	Address one year ago (international)
Yradpc	Address one year ago (postcode)
YrArr_year YrArr_mon	Arrival in the UK (year and month)

## Annex B: Eligibility for inclusion in item rate denominators

Table 1: Person questions

Eligibility criteria		Applicable variables
ALL PERSONS		SEX AGE MARSTAT SCADIND STUDENT
SCADIND = Yes, UK		SCADPC
SCADIND = Yes, Outside UK		SCADINT
SCADIND = Yes		SCADDTYP
STUDENT = Yes		TERMIND
TERMIND = This address or STUDENT = No	COUNTRY = WALES	COB  HEALTH CARER IDENTITY (IDENUK IDENINT) ETHNICID (ETHNIC ETHNIC1) WELSHLANG MAINLANG YRADIND DISABILITY PSSPRT PSPTL
	PSSPRT NOT = None	
	COB NOT UK (921-926)	ARRIVE_UK (YRARR_YEAR YRARR_MON) INTENTION
	ARRIVE_UK < 1 YEAR	
	MAINLANG NOT English	LANGPRF
	YRADIND = UK and AGE >0	YRADPC
	YRADIND = Outside UK and AGE > 0	YRADINT
	AGE >=16	QUALS ACTLW
	ACTLW = Working or EVERWORK = Yes	EMPSTAT SUPERVISOR
	ACTLW = Working	OCC_WORKING (Currently working) INDPUK_WORKING (Currently working) WORKADD (WKPLIND WKPLPC WKPLINT) TRANSPORT HOURS
ACTLW NOT = Working	EVERWORK LASTYRWORK	
EVERWORK = Yes	OCC_EVERWORKED (Worked in past) INDPUK_EVERWORKED (Worked in past)	
COMMUNAL RESIDENT		POSITION

Table 2: Household questions

Eligibility criteria		Applicable variable
All Households		TYPACCOM SELFCON ROOMS BEDROOMS CENHEAT
PERSON_COUNT >=1		TENURE CARSNO
	TENURE NOT= Owns	LANDLORD
RESIDENT_TYPE = 'Household' and PERSON_COUNT >=2		RE1PUK

Further reports in this series are [available to download](#)

Enquiries to Census Customer Services: [census.customerservices@ons.gsi.gov.uk](mailto:census.customerservices@ons.gsi.gov.uk)