

Overview of Administrative Comparator Data Used in 2011 Census Quality Assurance

November 2012

1. Introduction

The 2011 Census population estimates have been subject to a rigorous quality assurance process. As a part of this quality assurance process, a series of checks have been carried out that involve comparing census data to a range of other data sources, including administrative data. This paper gives an overview of these administrative sources with a focus on those used to quality assure census population estimates by age and sex.

The administrative sources covered are in the following sections:

- A [NHS Central Register](#)
- B [Birth Registration](#)
- C [School Census \(including data on independent schools\)](#)
- D [Customer Information System \(CIS\)](#)
- E [Higher Education Statistic Agency \(HESA\)](#)
- F [Super Older Persons database \(DWP\)](#)
- G [Child Benefit Claimant data \(HMRC\)](#)
- H [Migrant Worker Scan \(MWS\)](#)
- I [Assessors Portal](#)
- J [Armed Forces Data \(Defence Analysis Service Agency\)](#)
- K [Further Education Data \(Scottish Funding Council\)](#)
- L [Electoral Register](#)

Each of the sources is administered for different purposes and so inevitably covers different populations to those enumerated in the 2011 Census. This paper begins with an overview of the usual residence enumeration base for the 2011 Census.

2. 2011 Census enumeration base

A person's place of usual residence is generally the address in the UK at which they spend the majority of time. For most people this will be their permanent or family home. A usual resident of the UK, for census purposes, is anyone who, on 27 March 2011:

- Was in the UK and had stayed or intended to stay in the UK for a period of 6 months or more, or;
- Had a permanent UK address and was outside the UK and intended to be outside the UK for less than 6 months.

In addition to this, someone was counted as usually resident at an address if, on 27 March 2011 if:

- They were temporarily away from home, for example on holiday, visiting friends or relatives or travelling (unless outside the UK for 6 months or more).
- They were in a communal establishment such as a care homes or hospitals for less than six months.
- They were a baby born on or before 27 March 2011.

- They had more than one UK address and are staying at the second address on census night.
- They were present at the address, even if temporarily, and have no other usual address in the UK.

Someone was not counted as usually resident at an address on 27 March 2011 if:

- The address at which they were staying was not their usual address and they usually live elsewhere in the UK (these people would be counted as visitors).
- They were away from their home address and had been staying or were expected to stay in a communal establishment such as a care home or hospital for six months or more. In this case they were counted as usually resident at the communal establishment.

Further clarification was given for the usual residence definition of certain population sub-groups:

- **Students and children at boarding school** should be counted as usually resident at their term-time address.
- **Children with parents who live apart** should be counted as usually resident at the address they spend the majority of their time.
- **Armed forces** should be counted as usually resident at their permanent or family home even if the majority of their time is spent at their 'working' address.
- **People who live at more than one address** should be enumerated at their permanent or family home even if the majority of their time is spent at another address.
- **People with no usual address** should be enumerated at the address at which they are staying on 27 March 2011.
- **People in prison.** Convicted prisoners with a sentence of six months or more should be counted as usually resident in the prison. If the sentence is less than six months the prisoner should be counted at their permanent or family home. All prisoners on remand should be treated as visitors.

3. Administrative and alternative datasets use in quality assuring the 2011 Census

A. NHS Central Register

The NHS Central Register (NHSCR) data is a live database administered by National Records of Scotland (NRS). The database contains some 5.7 million individuals with approximately 50 - 60,000 new patient registrations (births) added each year as well as numbers of in- migrants and out- migrants. In the 2011 Census quality assurance process NHSCR data is used to assess:

- Age/sex estimates at Council area level.
- Age group/sex estimates at datazone level.

The NHSCR is the main source used in the quality assurance to check numbers of adults aged 16 to 64. Mid-year population estimates are also used (they are 'rolled forward' from the previous census).

Coverage

As a list of where individuals are registered, coverage only extends to those people who have registered with a GP. Examples of those groups not registered with an NHS GP include sole private care users, armed forces personnel, long-term prisoners, and patients in long stay hospitals. Individuals intending to stay in the United Kingdom for longer than three months can register with an NHS GP. This will mean that the register will include some individuals who would be excluded from the 2011 Census definition of usual residence (individuals who intended to stay for longer than six months). It is not possible to distinguish between short-term migrants (those who intend to stay for less than six months) and usual residents on the register, nor is it possible to identify whether all short-term migrants were removed from the register when they departed.

Accuracy

The accuracy of the NHSCR depends on all patients registering with an NHS GP when they move into or within Scotland. Accuracy of the register is known to vary across Scotland as individual Health Boards are responsible for its maintenance. List inflation exists on the register where individuals are no longer resident at the address at which they are on the register. Estimated list inflation is around 8 per cent on average for Health Boards in Scotland but is known to vary in some areas. It is acknowledged that the extent of any list inflation depends on the demographic profile of each area. Seven population sub-groups can be identified as being sources of list inflation:

- Students
- Pupils in residential schools
- Multi-occupancy dwellings
- Deceased patients
- Immigrants
- Patients who go abroad for more than three months
- Elderly patients over 100 years old

Timeliness

The extract of NHSCR data used by NRS for census quality assurance is taken with reference to census day. It is assumed that a six week lag should be allowed for individuals to change their GP registration, consistent with the assumed lag in patient NHSCR data used for official NRS estimates of internal migration.

B. Birth Registration

Birth registration data

Births that occur and are registered in Scotland are legally required to be registered. Registration is carried out by the Local Registration Service (LRS) in partnership with the National Records of Scotland (NRS). The extract used in census quality assurance includes all births which occurred in Scotland in the 12 months prior to census day.

In 2011, Census quality assurance, birth registration data is used:

- To assess age and sex estimates of babies under one year of age
- In fertility analysis where the number of births are used in conjunction with census estimates (and other comparators) to calculate fertility rates

Coverage

Birth registration is a legal requirement under the Births and Deaths Registration Act (1836), including those to mothers who are not usually resident. Not all births occurring in Scotland in the twelve months prior to census day will still be resident at census day. The register is not updated with babies who die or where migration takes place (either internationally or around Scotland). In 2010 for example, there were 218 infant deaths in Scotland

Accuracy

Information supplied at birth registration is generally believed to be correct since wilfully supplying false information may render the informant liable to prosecution for perjury. As noted information provided when the baby was born may not be the same as at census day when the extract refers to.

Timeliness

In Scotland, registration must take place within 21 days of the birth. The extract of data used in the quality assurance process is taken in May 2011 to cover the registration of all births which occurred in the 12 months prior to census.

C. School Census, Independent Schools data

The School Census is collected by the Scottish Government to ensure, amongst other things, appropriate school funding for each Council area in Scotland. There were approximately 670,000 children on the School Census when it was collected for the Census extract.

Independent Schools data for Scotland used in the quality assurance process also comes from the Scottish Government. Independent schools are those which are not maintained by Councils. Non-maintained special schools are not independent schools. The 150 independent schools cater for around 30,000 pupils and include day schools, boarding schools, single sex schools, co-educational schools, and schools catering for ethnic minority and faith groups.

In the 2011 Census quality assurance process School Census data are used to assess:

- Age/sex estimates for children aged 5-14 at local authority level (along with independent schools data)

Coverage

Parents of children of compulsory school age (5 to 15 at the start of the school year) are legally required to ensure that they receive a suitable education by regular attendance at school or otherwise. The School Census does not however cover children attending independent schools, pupil referral units and those who are home schooled. As a result approximately 4 per cent of children aged 5-14 are not covered, and this is particularly apparent in Edinburgh and Glasgow.

The extract of School Census data is based on a pupil's residence and so is consistent with the definition used in the 2011 Census. Independent Schools data however is based on the location of the establishment.

Accuracy

The geographic distribution of independent schools varies across Scotland. There are several council areas with fewer than 100 children at independent schools. In contrast, in Edinburgh there are around 10,000 pupils at independent schools. School Census data is based on age as at census day. As Independent Schools data is based only on age as at 1st December 2010 it was necessary to use basic assumptions on how many children will have reached their next birthday by census day.

Timeliness

The School Census and Independent Schools extracts used in the quality assurance process are based on the start of the autumn term (September 2010). This means that any migration between September 2010 and 27th March 2011 will not have been taken into account.

D. Customer Information System (CIS)

The Customer Information System (CIS) is the central repository of personal and biographic details for Department for Work and Pensions (DWP). It also includes certain information from HM Revenue and Customs (HMRC) with regard to the general working population. It also includes information about children whose parents have been awarded child benefit and individuals who are receipt of state pensions. Only aggregate CIS data is currently held by National records of Scotland (NRS).

In the 2011 Census quality assurance process CIS data is used to assess:

- Age/sex estimates for people of working age (16-64) at local authority level

Coverage

The CIS only covers individuals who have been allocated a National Insurance Number (NINo) either because they had Child Benefit paid for them or had applied for one as an adult (usually but not exclusively adult migrants). Some groups are not entitled to have a NINo if they do not have the right to work in the UK and do not have any entitlement to claim benefits including State Pension. This will include some international migrants who have retired as well as the spouses and other dependants of some migrants. Coverage is likely to vary across Scotland.

Accuracy

HMRC and DWP carry out extensive checks to ensure the accuracy, reliability, consistency, completeness and quality of the data. There may be lags in the updating of address information in some cases. For example Child Benefit claimant information is self-reported and receipt of the benefit (until age 16 years) is not dependent on correct address information.

Information on the CIS remains current until a death notification is received on to the system. Deaths registered outside the UK will not be automatically updated so individuals who have emigrated may well remain on the system. The number of individuals remaining on the system will reflect geographic variation in emigration patterns.

Timeliness

The CIS data used in the quality assurance process was extracted at census day.

E. Higher Education Statistics Agency (HESA) data on students in higher education

Higher Education Statistics Agency (HESA) data used in census quality assurance is based on data supplied to HESA by all publicly funded Higher Education Institutions (HEIs). Data is captured from each HEI's internal administrative systems which include student records, human resources and financial systems.

The HESA data is collected for the administrative purposes of the institutions but also the government departments and statutory bodies responsible for higher education policy and funding in the UK.

In the 2011 Census quality assurance process HESA data is used to assess:

- Estimates of students aged 18 years or over in full-time education (in conjunction with Further Education (FE) data).
- Estimates of students aged 18 or over in full time-education by accommodation type (residential or communal)

Coverage

The 2011 Census does not distinguish between higher education students and other individuals in full time-education such as those at school, at language schools or at further education institutions. In the QA process the census estimates of students only cover individuals who were aged 18 at the start of the academic year 2010-11. Similarly the HESA data used excludes individuals who were younger than 18 at the start of the academic year and individuals on part-time courses.

HESA data will include short-term migrant students who stay in the UK for less than six months. To be consistent with the census usual resident definition, some HESA records have been excluded where individuals are assumed to be short-term migrants. Students are assumed to be short-term migrants if they have arrived in the last six months and:

- i) Are exchange students with an expected length of study of a year or less (or not known) or
- ii) Are from non-European Union (EU) countries and have an expected length of study of a year or less.

HESA data used in the census quality assurance process is based on the student's term-time address. As the term-time address is collected at the start of the academic year, it is inevitable that in some cases term-time address will have changed by Census Day, March 27th. This is difficult to quantify but would be possible to assess with census records matched to HESA records. It was not possible to test this using 2001 Census data as HESA have only collected term-time address since the academic year 2007-08. Only individuals on higher education courses (post Higher) are covered by the HESA data used.

Accuracy

Not all student records have a term-time address available on the HESA data. The extent of this varies across the country between higher education institutions. In recent years between 5% and 10% of records were recorded without a term-time address. Term-time address has been imputed as part of the preparation for the census QA based on records at the same institution where a term-time address was provided.

Students who live in halls of residence are specifically flagged on the HESA data. Analysis of this indicated that in some instances this may not be updated when the student moves into residential accommodation.

Timeliness

The HESA data used refer to the academic year 2010/11. Students start and end date is used in combination to determine which individuals were still students on census day. Similarly, age at census day is calculated using the date of birth provided on the HESA data.

F. Super Older Persons database (DWP)

The Super Older Persons dataset from DWP contains information which has been derived from a variety of different data streams, such as TV Licensing, Winter Fuel Allowance, as well as Pension and Disability Allowances.

In the Census QA process, it is used to assess:

- the number of people aged 65 and over (by single year of age up to 89) for each Council area in Scotland; and
- the number of people aged 90 and over by single year of age (up to 100 years old) and the age group 'over 100' at Scotland level.

Coverage

An assessment of the coverage of this data since 2003 indicates that there is a close relationship between the mid year estimates and the data from this dataset.

The Super Older Persons dataset should include virtually all persons at and over the age of 65. This is partly due to the fact that some benefits and allowances due to older people are not means tested, and are therefore awarded universally, but also because of the wide range of benefits and allowances which are available.

It has been found that coverage was broadly consistent across all areas.

Accuracy

Address information and other demographic information for payment of benefits are known to be accurate and up-to-date. There are extensive checks in place to ensure eligibility and to prevent benefit fraud, and this ensures that the quality of information is good.

Timeliness

DWP derive this information on a quarterly cycle. DWP also allow a period of retrospection, during which data is checked and updated as necessary to ensure maximum accuracy and relevance. Census day (March 27th) lies between the reference dates used by DWP (the end of February and the end of May) and therefore data used for comparison with the Census was interpolated for comparative purposes.

G. Child Benefit Claimant data (HMRC)

Coverage

Take up of child benefit data for under-one year olds is estimated at 96 per cent. Take up increases with age but does not reach 100 per cent. This coverage is likely to vary across Scotland. Child benefit is subject to residence conditions and is, for example, only paid to parents of children who are ordinarily resident in the UK. As a consequence it is unlikely to cover short-term migrants.

Accuracy

Address information may not relate to the home address of the claimants. This situation may arise when parents and families have split to different locations. Child Benefit for children attending boarding schools may be claimed at the parental home.

Timeliness

Data obtained from HMRC for comparison with the Census was taken on Census day and is therefore fully compatible.

H. Migrant Worker Scan

The Migrant Worker Scan (MWS) is a subset of data from the National Insurance and PAYE System (NPS). It contains information on all overseas nationals (i.e. non-UK citizens) who have registered for and been allocated a National Insurance Number (NINo). Migrants need to register for a NINo to enable them to work, or claim benefits, within the UK. The MWS will only capture migrants aged over 15 years 9 months who wish to work legally or claim a benefit. It shows nationality at date of NINo allocation only so will not reflect any change in nationality after that date.

In the 2011 Census quality assurance process MWS data is used to assess:

- The number of international migrants who moved to Scotland from outside the UK in the twelve months prior to census day.

Coverage

The MWS does not include all adult migrants, only those who are non-UK citizens who have received a NINo in order to work legally or claim benefits. The extract also contains migrants who may not be residing in the UK. Some migrants may delay registering for a NINo for over a year after arriving in the UK or may never register. Moreover, applications for NINos by asylum seekers will not usually be approved until they have been granted rights to reside in the UK. Groups that are excluded from the analysis are:

- UK or dual-nationality migrants,
- Child migrants,
- Migrants who work illegally and
- Adult migrants who neither work nor claim benefits.

(This last group are commonly spouses of working migrants or students).

The MWS does not hold any information on how long migrants stay in the country. Some migrants may arrive in the UK, register for a NINo, and work for a short time before returning overseas. Similarly, a small number apply and are allocated a NINo before entering the UK and then never arrive usually because they are the dependent spouse of UK nationals resident abroad who are entitled to benefits normally Retirement Pension.

Accuracy

Residential address is recorded when a migrant registers for a NINo at Jobcentre Plus. Subsequent changes in address will only be recorded on NIRS-2, and hence the MWS, if the migrant or the Department for Work and Pensions (DWP) informs HM Revenue & Customs (HMRC) of the change. Anecdotal evidence suggests that DWP are most likely to be informed of address changes for migrants claiming benefits and HMRC for those self-employed and required to pay self employed National Insurance contributions and complete an annual tax return. It is likely that address information for some migrants will be out of date.

Assessment of the address information on the MWS indicates that some addresses are not residential e.g. bank, c/o, hotel. Workplace address may not be within the same local authority as residential address.

The Office for National Statistics (ONS) assessment of MWS published in 2009 found that a very small minority of records (less than 0.01 per cent) did not have any address information and around seven per cent did not have a postcode. DWP assign local authority geography using postcodes.

If no postcode was available, then post town and city was used. If no address information was available then records were randomly allocated to local authorities.

Timeliness

The extract of Migrant Worker Scan data used in census QA refers to March 27th 2011.

I. Assessors Portal - Council Tax data

Dwellings liable for council tax in Scotland are registered by local Assessors and information is available from the Assessors Portal. The extract used in census quality assurance includes all dwellings as at September 2010. In 2011 Census quality assurance, Assessors Portal data was used to assess the total number of dwellings.

Coverage

The Assessors Portal data includes all dwellings in Scotland. A particular issue with using this data is that it includes those dwellings that are unoccupied for the purposes of the census (e.g. holiday homes, second residences, vacant properties, and properties under construction). Council tax data provided directly from local authorities identifies such unoccupied dwellings - where the owner receives a discount if the address is not their main residence. Particular concentrations of second residences are found in the Highlands, Argyle and Bute and Eilean Siar. Furthermore, the census definition of a household and communal establishment may differ from the Assessors definition of a dwelling in some cases. For example, the census will list one university student hall as a communal establishment, whereas the Assessors Portal data may list each flat within the student hall as a separate dwelling. Again, local authority provided council tax data identifies where dwellings are part of a student communal establishment.

Accuracy

Information supplied by the Assessors Portal is generally believed to be accurate, as the local authority must inform the Assessors Portal when a property is extended or altered such that planning permission is required. However, there may be some cases where properties are converted or merged without legal planning permission, causing inaccuracies in the Assessors Portal data.

Timeliness

The extract from the Assessors Portal used in the quality assurance process is taken on or close as possible to Census night.

J. Armed Forces data (Defence Analytical Services Agency and United States Air Force)

Numbers of UK armed forces personnel used in the census QA process are based on data from the Defence Analytical Services Agency (DASA) and are taken from the Joint Personnel Administration (JPA) system. This data is also used in the annual production of mid-year population estimates.

In the 2011 Census quality assurance process armed forces data are used to assess:

- Estimates of armed forces personnel by age/sex for each local authority

Coverage

Data from DASA covers UK Regular Forces and includes all trained and untrained personnel. Gurkhas are included but full time reserve personnel, and mobilised reservists are excluded.

Accuracy

DASA data supplied refer to where personnel are stationed rather than where they are resident. A residence distribution has been applied to the DASA data to give greater comparability to the 2011 Census. The same approach is used as in the mid-year population estimates, applying a base-to-residence matrix compiled from information on both the workplace and residence of armed forces from the 2001 Census. This approach will be inaccurate not only because the distribution is ten years out of date but also because the instructions given to where armed forces personnel should consider themselves resident differed for the 2011 Census. In 2011 personnel were told to record themselves as resident at their permanent family home whereas in 2001 they were told to record themselves where they spent the majority of their time.

Timeliness

DASA data used refer to April 2011.

K. Further Education Data (Scottish Funding Council)

The number of individuals in full-time education at Further Education Institutions (FEIs) within Scotland is supplied by the Scottish Funding Council.

In the 2011 Census quality assurance process Further Education (FE) data are used (in combination with HESA data) to assess:

- Estimates of students age 18 years or over in full-time tertiary education

Coverage

The FE data used in the QA will only cover government-funded learners i.e. those funded by the Skills Funding Agency/Education Funding Agency. Information is not available for students in the further education sector who do not receive government funding.

This data does not include learners studying further education courses in schools or Higher Education Institutions however it does include those studying Higher Education (HE) courses at Further Education Institutions.

A wide variety of institutions are covered by the FE data used. These include general further education colleges (including tertiary colleges), special colleges (agriculture and horticulture colleges, art, design and performing arts colleges and specialist designated colleges), private providers (but government-funded learning) and other providers such as local authorities.

The data covers learners studying full-time courses. It does not include information on learners who are studying full-time for part of the year, those individuals on apprenticeships or on workplace or community learning courses.

Accuracy

Data is based on where individuals study rather than where they are resident. This will be inconsistent with the census where individuals will be at their term-time address. It is not possible to identify how many individuals study at an institution in a different local authority to the one in which they live. This is unlikely to be a major issue as most individuals in the FE sector attend their local institutions. Age has been calculated on March 27th and so will be consistent with age as recorded on the census form.

Timeliness

The data provided for Scotland refer to full-time learners enrolled on FE courses as at August / September 2010 only.

L. Electoral Register Data

The number of individuals who are on the Electoral Register is supplied by Electoral Registration Offices in Scotland. For Census QA this data is used to assess:

- Estimates of persons in the Census who are aged 18 and over

Coverage

The Electoral Register is constantly updated throughout the year. Address information is thought to be relatively good. There are some concerns over the coverage of the Electoral Register in areas of high population density, mobile populations and 'student' areas. There are also possible multiple records (mostly in cases of second homes and student registrations). Registration information is limited in demographic terms – no age information is collected other than that the person is 18 or over and there is no gender information collected.

Accuracy

The system of collection depends to some extent on 'notification of change' and therefore relies largely on information collected from previous canvasses. Electoral Registration forms are pre-printed and rely on respondents notifying any change. 'Household Registration' (where any individual in the household can provide the required information on behalf of the other residents) increases the potential for error. More than one canvass non response will prompt further investigation, and the registration is withdrawn, if no evidence of residence is found. An annual canvass is undertaken each year, where households are required to return forms, but a system of rolling registration is also in place, by which the Register is updated when new information or changed information is provided or obtained by Electoral Registration Officers.

Timeliness

For Census 2011 QA the Electoral Register in Scotland as at 1st April 2011 is used.