

Quality Assurance for Census

Quality Assurance of Administrative Datasets: NHSCR

**Dataset: National Health Service Central Register (NHSCR) - 30 June 2019
Extract**

Data Supplier:	National Records of Scotland (NHSCR)
Supplier info:	<p>National Records of Scotland (NRS) is a Non Ministerial Office of the Scottish Government. The purpose of NRS is to collect, preserve and produce information about Scotland's people and history and make it available to inform current and future generations.</p> <p>The NHSCR branch of NRS is responsible for maintaining the NHSCR, an electronic demographic database of all people born in Scotland, died in Scotland and those who have ever registered with a GP in Scotland.</p>
Data type	Unit records
Data content:	<p>The following variables are included at an individual record level:</p> <ul style="list-style-type: none"> • First name • Middle name • Last name • Previous names • Sex • Birthdate • Death date • Postcode • Date postcode was recorded • Posting (indicates which health board the person has registered to a GP in)
Time Period Covered	Extract as at 30 June 2019
Use of Data:	Quality assure methodology and processes in the 2019 Census Rehearsal

Data Source Information

The NHSCR is an electronic index for:

- every patient registered, now or in the past, with a Scottish general medical practitioner (GP),
- everyone born in Scotland since 30 September 1939, including those who have not been registered with a Scottish GP,
- patients formerly registered with a Scottish GP, who died after 29 September 1939.

The main purpose of the register is to permit the efficient movement of patient's medical record envelopes when they:

- transfer between Scottish Health Boards and health authorities in the rest of the UK;
- leave the country;

- join the Armed Forces (or are dependants of Armed Forces personnel).

The key inputs into the NHSCR are:

- Births (in Scotland)
- Deaths (from across the UK)
- GP Registration (within Scotland) – ‘migration’ into Scotland
- GP Registration (within the rest of the UK) – ‘migration’ out of Scotland

Data supply and communication

The data provided is done so annually under the terms of a data sharing agreement and includes record level data for a selection of variables as defined in a data sharing agreement for every person on the NHSCR.

The data is sent to the admin data team by the NHSCR team via approved NRS data transfer procedures as agreed in a data sharing agreement.

Quality Assurance undertaken by data supplier

The data entered by staff is regularly scrutinised. Supervisors check 5% of the work undertaken by staff each day to identify any potential training issues. These records are randomly selected based on subject matter, taking into account new areas of work, trends or concerns previously identified. This also helps the NHSCR to meet its service level agreement with the Scottish Government, NHS National Services Scotland which requires an accuracy level of 97%, which is currently being achieved.

As well as this, the NHSCR team undertake a variety of data quality initiatives on an annual/bi-annual basis where staff investigate the population of different variables in the register and to correct duplicates. These initiatives are carried out relatively frequently as they target areas of known concern and the findings are generally kept internal to the NHSCR team. These data quality initiatives include:

- investigating records where no death has been recorded for a person aged over 110 years old. In the majority of cases a death is traced (these are usually deaths that were missed at the time, usually from the 1970s or 1980s before the NHSCR was computerised) and the record is updated to reflect this.
- checking records where the postings variable is blank. This allows us to be confident that all records that should have a posting do. Where no posting exists it is usually for persons who are born in Scotland but they never registered with a Scottish NHS GP.
- populating records that do not have a Community Health Index (CHI) number¹ either with the CHI number if one exists or with a flag to show that there is not a CHI number for that record.

Extracts of the NHSCR are used by various statistical teams across the National Records of Scotland for a variety of purposes. NHSCR also collects feedback from these users of the NHSCR extracts where anomalies are identified and investigates these anomalies so a resolution or explanation can be found.

¹ <https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?ID=128&Title=CHI%20Number>

Quality Assurance undertaken by the admin data team within the NRS Census Programme

Once the admin data team receive the data, a number of data consistency and validation checks are performed, including:

- Checking the proportion of missing values for variables.
- Checking the validity of postcodes
- Checking the distribution of the population across different council areas and comparing this to previous years and/or existing population estimates.
- Checking the distribution of the day and month elements of dates of birth
- Checking the age distribution of the population.
- Checking that variables that should be unique are unique.

These checks are largely programmed with the output flagging up any anomalies, although analysts do also look at a small sample of records to spot any issues.

If these checks suggest the data may need to be amended/adjusted then the potential issues are communicated with the data supplier so the register can be amended if appropriate. However in this case these checks did not identify any issues with the data so this was not required.

Strengths and Limitations of the data source

Strengths	Limitations
<ul style="list-style-type: none"> • NHSCR is a comprehensive source of record level data that covers the vast majority of the population. • The data contains all of the variables used to link with other data sources (name, date of birth, postcode and sex) • births and deaths are added to the NHSCR database within 48 hours of them being registered with NRS 	<ul style="list-style-type: none"> • Generally does not include address information beyond postcode. There is a Unique Property Reference Number (UPRN) variable, however this variable is completed for less than 25 per cent of records. • It does not pick-up people who leave the UK (unless they informed their GP) leading to some inflation in the register • Moves within Scotland cannot be picked up until the patient registers with a new GP. As a result some people will be recorded in the wrong area. Particularly an issue among younger adult males. • There will be a lag in recent migrants into Scotland appearing on the NHSCR as they will only appear when registering with a GP. • Although a birth is added to the NHSCR within 48 hours, there can be delay in babies having a postcode (and posting) recorded until they are registered with a GP. • There is approximately a 2.5 month time difference between the NHSCR dataset and the Census Rehearsal 2019.

Risk/Profile Matrix

This section contains a risk/profile matrix for the NHSCR. The matrix reflects the levels of risk of data quality concerns and the public interest profile of the statistics. These have been determined by a review undertaken by the NRS Admin Data team using the information contained within the [Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit](#). For the use of data for this project, the cell highlighted is appropriate:

Level of risk of quality concerns	Public interest profile		
	Low	Medium	High
Low	Statistics of low quality concern and low public interest. [A1]	Statistics of low quality concern and medium public interest. [A1/A2]	Statistics of low quality concern and high public interest. [A1/A2]
Medium	Statistics of medium data quality concern and low public interest. [A1/A2]	Statistics of medium quality concern and medium public interest. [A2]	Statistics of medium quality concern and high public interest. [A2/A3]
High	Statistics of high data quality concern and low public interest. [A1/A2/A3]	Statistics of high quality concern and medium public interest. [A3]	Statistics of high quality concern and high public interest. [A3]

*A1/A2/A3 – definitions supplied [Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit](#).

Justification for Matrix Score

The public interest profile has been set to “medium” for the following reasons:

- There is a high level of interest in the Census Rehearsal 2019 and this work provides important information to assess the quality assurance methods and processes to be used in the 2022 Census. Being able to assess these methods and processes accurately will therefore improve the quality of the 2022 Census which is used for a wide range of purposes, including planning, allocating funding and informing policy.
- However, as the Census Rehearsal does not produce statistical outputs that will be used outside of improving Census methods and processes we would not consider it of ‘high’ public interest. However, it does support evaluation

reports and methodology papers which will be published on the Census website, these are of more interest to Census users than the wider public.

The risk of quality concerns has been set to “low” for the following reasons:

- These are inevitable issues given the nature of the data collection and cannot be avoided by the supplier such as inflation when people don't update details if leaving Scotland and lag in recent migrants appearing on the register. However as these are known issues they can be accounted for when using the data.
- The risk of quality concerns is reduced due to the service level agreement to have at least 97% accuracy that is being met.
- This is further reduced as the NHSCR team have a variety of data quality initiatives that are undertaken on a regular basis to mitigate these data quality issues.
- The NHSCR team and the Census programme both fall in the Statistical Services division of NRS and both report to the same Director. This means that there is an increased awareness of issues each other may be facing and the impact this may have on the other party. We can therefore be confident that we will be made aware of any changes that would have an impact on how this data is used.
- As this data was extracted as of June 2019, there has been no impact to quality due to COVID-19.