Office for National Statistics

# NHSCR data: quality assurance of administrative data used in population statistics, Dec 2016

The quality assurance undertaken on administrative data for NHSCR data used within Population Statistics Division (PSD) publications.

Contact:
Pete Large
pop.info@ons.gsi.gov.uk

Release date:
20 December 2016

Next release:
To be announced

## Table of contents

# 1 . Introduction

Internal migration – that is, moves between areas within the UK – is an important part of population change in local areas. We produce estimates of internal migration between local authorities in England and Wales and between other parts of the UK and local authorities in England and Wales. These estimates are important as they feed into the main mid-year population estimates for local authorities in England and Wales, and into the national and subnational population projections we publish.

As there is no comprehensive system within the UK to record movements of the population, we use a range of administrative data sources to estimate internal migration. These are: the National Health Service Central Register (NHSCR) (as it appears in each part of the UK), the Patient Register (PR); Higher Education Statistics Agency (HESA) data; and the Northern Irish medical cards register. We use this combination of sources because each helps overcome coverage or quality limitations in the others.

This report covers the NHSCR data. This data is used in estimating internal migration in 2 ways:

- estimating the number of migration moves where the person is not present on the main Patient Register at the start, or the end, of the year (for example, a baby born during the year who moves before the end of the year)
- estimating “cross-border flows” between England and Wales and the other parts of the UK

Similar reports are being published on the other administrative data sources used. Our Population Estimates for the UK, England and Wales, Scotland and Northern Ireland release reproduces estimates of internal migration produced and published by National Records for Scotland. These NRS estimates also use the Community Health Index data source to disaggregate estimates to the level of Scottish Council Areas. The quality of this data for this use is discussed in the NRS report Information on the Quality Assurance Arrangements for Administrative Data used in population estimates. Data for Northern Ireland is not covered by this source, a separate QAAD report will be published on this data in early 2017.

This report covers the processes, from data collection through to the statistics published by our Population Statistics Division (PSD), which use this NHSCR data as an administrative data source. This report is designed to help people better understand the level of quality assurance applied to this data prior to use by PSD. It identifies potential risks in data quality and accuracy as well as details of how those risks are mitigated.

This report does not aim to report on the whole of the process of estimating internal migration. In addition to the other data sources mentioned above (for which QAAD reports will be published no later than early 2017), further information relating to the quality of the internal migration estimates can be found in the Internal migration methodology document, the Internal migration QMI and other guidance and methodology documents.

Within Population Statistics Division, NHSCR patient data were assessed separately by relevant teams using the UK Statistics Authority’s Quality Assurance of Administrative Data Toolkit. The results of those assessments are that NHSCR data received an A2 rating for England and Wales, and A1 rating for Scotland information. The scores provided by each team and the rationale behind those scores will be provided, in terms of both the risk and profile components, later in this report.

**Summary of Quality Assurance Level for NHS Central Register**

| Use | Risk of quality concern | Public interest | Assurance Level |
|---|---|---|---|
| Overall Assessment | | | A2 – enhanced |
| Internal migration England and Wales | Low | Medium | A2 – enhanced |
| Cross-border flows | Low | Medium | A1 – basic |

Source: Office for National Statistics

## Quality Assurance Level for Internal migration: A2 – enhanced assurance

## Level of risk of data quality concerns: Low

- The NHS Central Register is used to account for additional moves that cannot be identified on the NHS Patient Register.
- Quality and statistical bias issues are understood and measures have been taken to mitigate these.
- Service Level Agreement in place to assure standard level of data received.

## Public interest profile of the statistics: Medium

- Internal migration estimates are published National Statistics.
- Though the internal migration estimates themselves are relatively low-profile they feed into the mid-year population estimates and projections, which are National Statistics which attract media interest and have a wide variety of other onward uses including resource allocation.

If you feel that this report does not adequately provide this assurance then please contact pop.info@ons.gov.uk with your concerns.

The toolkit outlines 4 areas for assurance; the rest of this report will be split into these areas, with further subdivisions by country. The areas for assurance are:

- operational context and administrative data collection
- communication with data supply partners
- QA principles, standards and checks applied by data suppliers
- Producer's QA investigations and documentation

# 2 . Data collection

The National Health Service Central Register (NHSCR) began life as the National Register, which was used to issue identity cards and ration books and to help with the call-up for the Armed Forces in WW2. When the NHS was set up in 1948, the National Registration numbers were used to ensure that each patient's record had a unique identification. After rationing ended, the register was retained as the Central Register for the NHS for England and Wales. This ensured that Primary Care Support Services (PCSS)[1] maintained up to date lists of patients resident in their areas. The NHSCR at Southport provided a comprehensive system to assist with NHS patient administration in England and Wales. Similar, but separate, systems are maintained in other parts of the UK by NHS Wales Information Shared Services, National Records for Scotland and the Northern Ireland Central Services Agency for the Health and Social Services. NHS Digital and National Records of Scotland are able to provide us with the required information from all these systems.

The NHSCR covers a range of services; one of these is to record the transfer of patients between PCSS. This data is collected and used as a proxy for internal migration estimates. Each record in the register contains the NHS number, name, date of birth and date of acceptance by the PCSS. The entries on the register are updated on receipt of information from PCSS. If the updating involves a change of PCSS (because the GP falls within a different PCSS) a "migration" record is created. In addition to patient moves, migration records are created for those emigrating and later returning, and for new immigrants from outside the UK. Some patient moves do not count as migrations such as prisoners, mental health patients and armed forces.

The NHSCR system processes around 2 million records of movement per year to produce statistics on population movements within the UK. Individual records for moves within England and Wales, and moves into England and Wales from Scotland, Northern Ireland and abroad are extracted from the NHSCR and sent weekly to us for processing.

The registration process begins when a patient attends a GP surgery and asks to register with a GP[2]. The patient completes a registration form (Form GMS1 – Registration with a GP) and may be asked for their medical card, this may not be available depending on local policy on issuance of medical cards on transfer from one area to another. Further information on date of birth, sex, current address and postcode, name of their last GP, previous address and postcode and whether the patient has come from abroad will be requested. In April 2006 this process was changed to a welcome letter being sent to the first registration only, that is, not to transfers. This is likely to reduce the availability of medical card type information over time.

GP practices notify the PCSS of new patients. These are processed by the PCSS with the patient records being checked to see whether the person is already present on the local database. This may result in a delay in the registration which could result in the person being reported as moving in the incorrect data period. If the person is new to the area, NHSCR are informed so that medical records can be obtained. Patient registration process may vary between England and Wales GP practices.

NHSCR maintains a central index, which holds a unique entry for each patient. Information held includes the NHS number, forename and surname, date of birth, sex and health authorities' (HAs) code (of current and previous HAs). Unlike the individual registers held by PCSS, the postcode or address is not available. When a patient registers with a GP practice the practice will notify the relevant PCSS of the registration. The PCSS establishes whether the patient is already registered within the area. The PCSS will notify the NHSCR of any new patients in its area. The NHSCR can also be updated following a birth or when someone emigrates. NHS Digital supply us with an extract of this data.

When a patient registers with a GP practice the practice will notify the relevant PCSS of the registration. The PCSS establishes whether the patient is already registered within the area. The PCSS will notify the NHSCR of any new patients in its area. The NHSCR can also be updated following a birth or when someone emigrates. NHS Digital supply us with an extract of this data.

Notes for Data collection:

1. The generic term of Primary Care Support Services (PCSS) covers current and historic NHS organisations such as Family Health Services Authorities, Primary Care Trusts, and Local Health Boards in Wales.

2. For simplicity, the description of processes here focuses on the process in England rather than covering each NHSCR system separately.

# 3 . Communication with data supply partners

Communication with the data supply partners NHS Digital and National Records for Scotland is managed by the Internal Migration team within our Population Statistics Division (PSD).

## 3.1 NHS Digital

The majority of NHSCR data is supplied to us by NHS Digital. Supply of this data is documented in the Service Level Agreement (SLA) between The Statistics Board [ONS] and The Health and Social Care Information Centre [NHS Digital] for the provision of data and advice under Annex (2a). This SLA forms part of a MOU (Memorandum of Understanding) with Department of Health and NHS Digital (formerly HSCIC).

This SLA covers the provision of the following services:

- covers the provision of migrant data to ONS
- ensures that data quality is monitored and performance is reported on
- ensure communications between both parties
- details the procedures for issue resolution and escalation
- sets out procedures for handling system and processing changes
- outlines data security provisions
- specifies the frequency of data supply

Should the SLA require modification there are procedures within it which sets out how ONS and NHS Digital are to agree the modifications. This agreement will normally be reviewed annually in June by the Steering Group, at the same time as the annual reports, and reissued, incorporating any changes recorded in the Record of Amendments.

The main formal liaison is through the ONS/NHS Digital Steering Group comprising of the SLA Managers and Annex Managers from the NHS Digital and ONS. The Steering Group meetings are held at least annually or more frequently as required, to review progress and monitor performance against the SLA. Any changes to the provision of the services specified in the SLA are signed off by the Steering Group.

Change management is via discussion between us and NHS Digital and subject to agreement in writing. As part of the regular process of supplier engagement the working level group for each Annex will meet to discuss any changes that affect the data and checks whether the metadata and Data Supply Template remain the same, or need updating. These groups will meet as necessary and will be chaired by the appropriate ONS Annex Manager.

Data exchanged in the services described in the SLA between ONS and NHS Digital is carried out under the legal cover given by section 43 of the Statistics and Registration Services Act 2007 for supply of patient data relating to NHS registration for processing by us as a service in support of the production of population statistics.

There are 2 types of data supply provided under this agreement, corresponding with the two uses made of the data in estimating internal migration.

i) Estimating the number of migration moves where the person is not present on the main Patient Register at the start, or the end, of the year NHS Digital provides us with weekly extracts of migrant moves referring to the previous week and a weekly file showing the number of “no trace rejections” identified by NHS Digital support team. We monitor these weekly and investigated if the total number of “no traces” is significantly higher than usual. NHS Digital provides an incident management process to resolve any data quality issues we identify.

The data is supplied weekly via email as a fixed width format file with the following variables:

- record type (this provides an indication of the location within the dataset: first record, last record or body record)
- sex
- date of birth
- health authority of origin
- health authority of destination
- date of acceptance (this represents when the NHSCR was last updated)

In order to protect anonymity, NHS number and name are not requested, geographical information is at the level of grouped local authorities to help further reduce possible identification.

The data is held on a server as a flat file without encryption and access is restricted to named persons using standard active directory restrictions. The file containing a weekly data extract is emailed to our internal migration team.

ii) Estimating “cross-border flows” between England and Wales and the other parts of the UK NHS Digital supplies us with an annual file of flows to and from Scotland and Northern Ireland to enable us to produce internal migration estimates.

The data is supplied annually as a fixed width format file with the following variables:

- NHS number
- old NHS number (this may be used due to a change in NHS numbering system)
- date of birth
- date of acceptance (this represents when the NHSCR was last updated)
- country code of origin (migration indicator)

The data is held on a server as a flat file without encryption and access is restricted to named persons using standard active directory restrictions.

The file containing the annual data extract is emailed to a secure NHS mail account and accessed by a member of our Internal Migration team. Again, NHS Digital provides an incident management process to resolve any data quality issues we identify.

## 3.2 National Records of Scotland

An additional data supply is provided by National Records of Scotland (NRS). This supply is governed by a Data Access Agreement. Data is supplied quarterly in 2 files:

Quarterly NRS File 1 – Text file from NRS containing the variables Sex, Date of birth, Health Authorities of origin /destination and date of processing:

- moves from England and Wales to Scotland

Quarterly NRS File 2 – Text file from NRS containing the variables Sex, Age (year only), destination Health Authority and date of processing for moves to Scotland from Northern Ireland:

- moves to Scotland from England and Wales
- moves to Scotland from Northern Ireland

The data is held on a server as a flat file without encryption and access is restricted to named persons using standard active directory restrictions.

These files are emailed to the member of the ONS Internal Migration team who requested the data.

## 3.3 Engagement with users

Population Statistics Division (PSD) continually engages with users to understand how well outputs meet their requirements. PSD's user engagement activities include formal consultations on proposed changes to outputs, regular communication on plans through a quarterly newsletter, and external events open to all users. In addition, where evaluating changes to methods or sources has required specialist knowledge of local areas, PSD has organised Local Insight Reference Panels to elicit the views of relevant local authorities. From these activities, any issues relating to the sources, and their fitness for the proposed use, will naturally come out. Issues restricted to one output will generally be addressed by the team responsible for that output while the Stakeholder Engagement team in PSD takes an overview of any issues with more general implications, and ensures that this is considered in development of outputs across the division. It should be noted that users are more likely to comment on the overall methodology and the effect that it has on the final statistics than on a contributory data source.

Any issues around the quality of the statistics are described in the Quality and Methodology Information report accompanying each output. Issues around specific administrative data sources used in producing the statistics are considered in Quality Assurance of Administrative Data reports such as this.

When changes are proposed to methods (including changes in data sources being used in producing statistics) the ONS Population Methodology and Statistical Infrastructure Division will assess the resultant methods prior to implementation to assure that they are of sufficient statistical quality to meet user needs and are an improvement on the previous method. An independent evaluation by academic experts may also be undertaken, should methodological changes be extensive. The methods are also subject to scrutiny by the UK Statistics Authority as part of the National Statistics accreditation programme under Principle 4 of the Code of Practice for Official Statistics (sound methods and assured quality).

The Responsible Statistician is named for each release and contact details for them are provided, so should someone have concerns over the statistics they are able to communicate them. Methodology documents are published to enable users to provide scrutiny.

# 4 . Quality Assurance by the data suppliers

This section details the checks and standards applied to the data prior to receipt by Population Statistics Division [1] . The checks carried out by Population Statistics Division upon receipt of the data are detailed in Section 5.

## Data quality

It is essential that the quality of the data extracted from the National Health Service Central Register (NHSCR) is of a high quality. The data are used to identify migrants, where they moved from and to, and details of their age and gender. If the quality of these variables is poor, the resulting migration estimates will be of very little value.

The NHSCR system has built in database integrity checks to ensure the continued accuracy of the data. In addition, major database integrity checks may be run monthly by the GP practices and checked by support teams to identify and rectify any problems.

The current NHSCR data have other issues that impact statistics (recording only moves between NHAIS areas) so a combined method using both sources is required.

## Accuracy of variables on the NHSCR

NHSCR data are collected and processed weekly. The data held are the patient's NHS number, name, and date of birth, health authority where the patient is or has registered, along with acceptance dates for these health authorities. There are several data quality issues surrounding the NHSCR data. Firstly, there is a time lag between a person moving and the NHSCR being notified of that move. This time lag can vary from one area to another. Secondly, the NHSCR includes a number of migrants who have moved more than once during the year.

## Patients not recognised on the NHSCR

In some cases, a person may not be found on the NHSCR. In such cases, the person is allocated a new NHS number. A migration record is prepared with only the destination health authority code. The code for the origin health authority is left blank. Following computerisation, the proportion of persons that cannot be traced has reduced considerably.

## NHS Number allocation capability

Prior to October 2002, NHS Numbers were issued, by Registrars of Births and Deaths at the point of civil registration, to children born in England, Wales and Isle of Man. In 2002 this process changed and NHS Numbers were issued to new born babies at Maternity Units or by Child Health. This process change had no discernible impact on the quality of data being passed to us from NHS Digital.

In 2006 the capability to allocate NHS Numbers was given to National Health Applications and Infrastructure Services (NHAIS). This was the first stage in widening the allocation capability for non-births to approved and compliant NHS organisations. Increasingly, the NHS Number allocation capability has passed to a wide range of NHS organisations at secondary as well as primary care. Note that, an allocation made at primary care is accompanied by a Primary Health Care (PHC) posting, that is, the patient is registered with a GP and receives a NHAIS posting. Allocations made in secondary care are not accompanied by a posting.

## Registration v Residence

Prior to 2012, patients in England and Wales were registered based on their residence, irrespective of the location of their GP. This resulted in the concept of “fringe” patients, that is, those that are resident in an area different from the area of their GP. From 1 April 2012, patients in England only were able to register with their GP of choice irrespective of their residence. From this point in time new PHC registrations were based on the GP registration area rather than the area of the patient residence (England only – patients in Wales are still registered based on their residence).

Whilst the majority of new PHC registrations will continue to be with a local GP, the posting that will be sent to us by CHRIS will represent the location of the GP and not the location of the patient. This may mean that a patient could now change GP but not move house and this could generate a migration notification. Conversely, a patient may move house from one geographical area to another but not change their GP and this would not generate a migration notification.

Since April 2012, there has been a programme of work to migrate fringe patients from the NHAIS of residence to the NHAIS of GP registration. These have been administrative changes and, as with the transition of previous codes to DMS a migration request was not generated.

Note that the transition of fringe patients from one NHAIS to another remains incomplete without a definite timescale for completion.

## Reduced data received by CHRIS.

The NHS no longer records patients with the following postings:

- Diplomatic Embarkation
- Prisoner
- Psychiatric patient

This has been the case since 2008, when all postings of these types were changed to Cancellation.

Additionally, it used to be common practice for patients who were leaving the country either permanently or for extended periods to relinquish their medical cards at point of departure. These were routinely returned to the NHS Digital and the corresponding patient record updated with an embarkation posting. With the passing of time fewer and fewer patients handed in their medical cards, and in some cases patients did not have a medical card available to hand in due to the policy of their local area not to issue such a card. The consequence of these process changes is that there will have been some small and gradual degradation of data quality.

NHS Digital provides a range of data services which include developing and assuring national systems against appropriate contractual, clinical safety and information standards. Further information can be found on the NHS website.

The patient lists are maintained by the Primary Care Support Services (PCSS) following guidelines for the removal of patients who have died, left the country, no longer live in the PCSS area or no longer known at that address.

To ensure lists are kept up to date NHS England has put in place a rolling list maintenance programme which include the following:

- routine business processes
- routine list inflation work
- multiple occupancy checks (monthly)
- university/college student/residential school checks (annually October to December)
- patients aged 100 and over (six monthly March and September)
- immigrant checks (monthly)
- notification of demolished addresses (quarterly)
- patients not seen by general practice in previous five years (annually)
- additional measures – comparing figures with our mid-year estimates etc
- monitoring

Further information can be found on the NHS England website in the document ' Tackling list inflation for primary medical services'.

We are currently researching how the use of data from emerging new data sources within the health system might provide better quality data in the future for producing internal migration information.

The NHSCR does not provide sufficient geographic detail for use as the sole NHS source. It does provide moves throughout the year and can be used in the internal migration estimates to scale up the number of moves for various groups not identified on the Patient Register:

- moves of those not in the UK or not born at the start of the year
- moves of those who have left the UK or died by the end of the year
- interim moves of people who have moved more than once during the year

Notes for Quality Assurance by the data suppliers:

1. Again, for brevity, the description of processes here is primarily restricted to the NHSCR system for England.

# 5 . Quality Assurance by ONS

Different Quality Assurance (QA) processes are applied for the three data supplies as described below:

## 5.1 England and Wales: Weekly NHSCR files

Weekly National Health Service Central Register (NHSCR) Files – The variables are read in and derived variables are created. The variables are then validated and missing values are imputed before being loaded into the live data sets.

The NHSCR weekly data is processed and the first check is done manually on the number of records received and the total validated after being loaded into the system. The system then automatically carries out validation checks and produces 3 reports:

- Validation report
- Error report
- Imputation report

The validation checks are:

- age must be greater than zero and less than 125
- sex must be male or female
- the area of origin cannot be the same as the area of destination

In most cases, if values are missing or invalid, they are imputed. Sex will be set to male or female alternately and age will be imputed based on the age distribution of the rest of the dataset. However, if either the origin or destination area is missing or invalid, the record is excluded. The validation reports are checked and any issues with the data will be investigated before being reported to NHS Digital.

## 5.2 Scotland: Quarterly NHSCR files

Quarterly NRS File 1 – Text file from National Records of Scotland (NRS) containing the variables Sex, Date of birth, Health Authorities of origin or destination and date of processing:

- moves from England and Wales to Scotland

Quarterly NRS File 2 – Text file from NRS containing the variables Sex, Age (year only), destination Health Authority and date of processing for moves to Scotland from Northern Ireland:

- moves to Scotland from England and Wales
- moves to Scotland from Northern Ireland

The NRS quarterly data is loaded into the system and the first check is done manually on the number of records received and the total validated after being loaded into the system. The system then automatically carries out validation checks and produces 3 reports:

- Validation report
- Error report
- Imputation report

The validation checks are:

- age must be greater than zero and less than 125
- sex must be male or female
- the area of origin cannot be the same as the area of destination

In most cases, if values are missing or invalid, they are imputed. Sex will be set to male or female alternately and age will be imputed based on the age distribution of the rest of the dataset. However, if either the origin or destination area is missing or invalid, the record is excluded. The validation reports are checked and any issues with the data will be investigated before being reported to NRS.

Further information on processing and quality assurance is covered in the internal migration desk instructions.

### 5.3 Cross-Border Flows: Annual NHSCR file

The data received annually from NHS Digital is input into the internal migration live system.

The data is validated by running a program which checks the following:

a. Missing NHS numbers

b. NHS number is not equal to a length of ten characters

c. Missing dates of birth

d. Invalid date of birth

e. Missing acceptance dates

f. Invalid acceptance date

e. Missing Migration indicator

g. Invalid Migration indicator (not Scotland or Northern Ireland)

If any instances of these occur, the records are sent to a “Dirty” file and all other records are sent to a “Clean” file. This process includes validation checks to identify missing and invalid entries which are investigated and if above the agreed threshold are reported to the data supplier to get the data re-supplied. Counts are also compared with previous years to ensure the difference is acceptable.

Further quality checks are completed on the final output files for the variables, age breakdown and counts and same scales as previous years and these differences are not unreasonable.

## 6 . Conclusion

## 6.1 Strengths and limitations

Internal migration is difficult to estimate as there is no compulsory system within the UK to record movements of the population. Data from the NHS are seen as offering the best proxy for internal migration moves because most people are registered with an NHS GP and most people will notify their GP or reregister with a new GP when they move. A combination of administrative data sources is used to help overcome coverage or quality limitations.

The National Health Service Central Register (NHSCR) provides details of patient record changes based on GP registrations, between Primary Care Support Services (PCSS). These are usually made up of groups of local authorities. It includes almost everyone resident in England and Wales and because the NHSCR data is updated weekly, multiple moves during the year are included. It also records migrant babies aged less than one year, who would not be on a register at the start of the year and would otherwise be missed.

Therefore, the coverage of the general population by the NHS, the effectiveness of the administrative process, the quality of the variables held by the NHS and the extent to which people re-register with a doctor when they change their address will affect the accuracy of the internal migration estimates derived from the NHSCR.

There are known limitations using the NHSCR data and these include the following:

- the data only being available for PCSS and not at local authority level
- the data exclude moves of people who delay re-registering with a GP after moving – referred to as “lagging”. This means that their move may be estimated in a later year than when they actually moved. However, the impact on the internal migration estimates themselves may be limited: any moves missed because of lagging are likely to be offset (to a greater or lesser extent) by previous “laggers” now re-registering
- the data exclude moves of people who never re-register with a GP after moving (for example, if they emigrate or die)
- the data cannot be used to identify moves for specific reasons, in particular moves either in or out of prison or the armed forces
- the data does not cover moves by those people that only use private medical services; this number is small, particularly as most private patients will also be resident on NHS registers