Internal Migration Estimates

Introduction

ONS is responsible for producing estimates of internal migration in England and Wales. However, migration is recognised as the most difficult component of population change to estimate as there is no compulsory system within the UK to record movements of the population. At present ONS uses a combination of three administrative data sources as a proxy for internal migration within England and Wales; namely the National Health Service Central Register (NHSCR), the Patient Register Data Service (PRDS) and Higher Education Statistics Agency (HESA) data.

ONS publish rolling-year interregional migration estimates on a quarterly basis. These estimates are based solely on NHSCR data. These estimates are timely (published around 8 months after the reference period), but are only available at a higher level of geography; namely the regional and Former Health Service Authority (FHSA) level.

Annual mid-year estimates of internal migration are based on a combination of NHSCR, PRDS and HESA data and are available at local authority level. ONS considers the annual mid-year estimates consisting of a combination of administrative data to be more complete and it is these estimates which are used in the calculation of mid-year population estimates.

2. Quarterly rolling-year interregional migration estimates

This section covers the quarterly rolling-year interregional migration estimates looking at publications, data sources and how the estimates are calculated.

2a. Publications

Quarterly rolling-year interregional migration estimates are produced using NHSCR data. Four tables are published on the ONS website each quarter (February, May, August and November):
Internal Migration: Interregional migration within the UK
Internal Migration: Flows by sex and age group
Internal Migration: Square matrix showing origin and destination
Internal Migration: Square matrix showing origin and destination by age and region

2b. Data source

The NHSCR covers a range of services; one of these is to record the transfer of patients between FHSAs. These data are 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 FHSA. The entries on the register are updated on receipt of information from FHSA. If the updating involves a change of FHSA (because the GP falls within a different FHSA) a 'migration' record is created. In addition to patient moves, migration records are also created when a patient removes himself from the NHS doctor's list to enter the armed forces, and then again when they return to a civilian NHS doctor. Finally, migration records are created for those emigrating and later returning, and for new immigrants from outside the UK. However, some patient moves do not count as migrations. For example, moves to long-stay psychiatric hospitals and imprisonment. In addition, internal migration estimates do not include movements of the armed forces.

The NHSCR migration system processes around two million records 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 ONS for processing.

Moves into Scotland from England, Wales and Northern Ireland are extracted from data held by the National Records of Scotland (NRS) and sent quarterly to ONS for processing. Data on flows into Northern Ireland are received by ONS from NISRA in the form of quarterly tables.

2c. How the estimates are calculated

This section covers the methdology behind the rolling-year interregional migration estimates and looks at the following; validation and imputation, creation of derived variables and cross border flows.

Validation and imputation:
Data are received weekly from the NHSCR and quarterly from Scotland and Northern Ireland. 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 validation of the weekly NHSCR data files includes the following checks:
-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.

Creation of derived variables:
The date of birth variable is used to derive age at migration. Date of migration is calculated as the acceptance date at the new FHSA minus 30 days for England and Wales or 60 days for Scotland.

Cross border flows:
Information on moves into Scotland and Northern Ireland are collected and treated differently from moves within England and Wales.
-Data on moves from England and Wales to Northern Ireland are received from the NHSCR.
-ONS receive data sheets from NRS stating the number of moves to and from Scotland to and from England, Wales and Northern Ireland.
-ONS receive quarterly totals of the number of migrants to Northern Ireland from England, Wales and Scotland.

Invariably, the number of migrants moving to Northern Ireland as recorded in the NHSCR is different from the number of moves to Northern Ireland recorded by NISRA. The consensus is that the country that receives the migrants is likely to have a more accurate count than the country of origin. Therefore, ONS apportions the NHSCR data to take account of the differences in the number of moves recorded by the NHSCR and NISRA.

3. Annual mid-year internal migration estimates

This section looks at the annual mid-year internal migration estimates and covers publications, data sources and how the estimates are calculated.

3a. Publications

Annual mid-year internal migration estimates are produced using a combination of PRDS, NHSCR and HESA data. These estimates are available at local authority level. Four tables are published on the ONS website each year:
Table 1: Numbers to and from each local authority in England and Wales from and to the rest of the UK

Table 2a: Square matrix of local authority level moves

Table 3: Square matrix of local authority level moves within each region and to each region by age

Table 4: Numbers to and from each local authority in England and Wales from and to England and Wales.

The annual release is accompanied by a statistical bulletin which explores the data in more detail.

3b. Data source

Three data sources are used in the calculation of annual mid-year internal migration estimates; PRDS, NHSCR (covered in section 2) and HESA.

PRDS:

PRDS is the main component of the mid-year internal migration estimates. Every Primary Care Trust (PCT) in England and Wales holds patient registers. The patient registers contain information on the patients registered with NHS GP within their area of responsibility. By obtaining a download from each patient register on an annual basis and by combining all patient register extracts together, it is possible to create a total register for the whole of England and Wales. Comparing records in one year with those of the previous year by linking on NHS number enables identification of people who change their postcode during the year. A migrant is defined as a person who, between one year and the next, changes their area of residence. This final step leads to an estimate of the number of people who have moved during the year, from the number who have changed their postcode following a move.

The data are downloaded from the patient registers at each health authority as at the end of July each year to enable migration estimates to be made for the year ending 30th June each year. This is consistent with the assumption that people delay registering with a new GP by approximately a month after they move. The variables available to ONS are shown in Table 1:

Table 1: Variables on the Patient Register data

Variable	Details
NHS number	
Date of birth	
Sex	The negligible number of 'indeterminate sex' cases are randomly allocated a gender.
Postcode	Postcode of current residence
Date of acceptance	Date first registered in health authority
Patient's health authority	Health authority supplying the data – used for validation purposes
GP's health authority	Can differ if patient lives outside the catchment area – used for validation purposes
Registration type	Used to identify births, first acceptances, transfers from another health authority, immigrants, ex-service – used for validation purposes
Previous health authority	This is not a reliable variable as it can be overwritten with a GP id – used for validation purposes

Migration estimates derived from solely from PRDS miss some groups of migrants. By comparing patient registers in two consecutive years, certain groups of moves that occur during the year will be missed. This occurs because it cannot capture the movement of those migrants who for one reason or another were registered with a doctor in one of the two years, but who moved during the year. The largest group of these is migrants babies aged less than one year, who would not be on the register at the start of the year.

Other people who are not on the register at the start of the year, but who move after joining the NHS and before the end of the year, would also not be captured. Such people could include:
-those leaving the armed forces
-international immigrants.

Similarly, people who move within the year, but are not on a register at the end of the year would also not be captured. Such people would include anyone who moved and then:
-before the end of the year, either died or enlisted in the armed forces
-emigrated from the country.

Another key point is that due to its point in time nature, the PRDS will fail to capture people who move multiple times within the year, whereas all within year moves are included in the migration estimates derived from the NHSCR. Therefore to ensure the most accurate coverage possible, the more complete information from the NHSCR is combined with the more geographically detailed

data from the patient registers. More details on the constraining processing are available later in the document.

HESA:

One of the known limitations of relying on GP registration data is that young people, particularly young men, can be slow to change their registration when they move. One of the most common reasons for migration among young people is to attend a course at a higher education establishment, so this limitation of the current internal migration estimation process is a key issue in the estimation of internal migration for this population sub-group. In May 2010, ONS introduced an additional adjustment for students based on HESA data to improve this undercount. HESA data contains records for students registered at higher education establishments and includes both term-time and domicile address variables. An adjustment is made to both first year students moving to higher education establishments and moves made by students at the end of their studies. For more information on the student adjustment methodology please see the ONS website (http://www.ons.gov.uk/ons/guide-method/method-quality/imps/msi-programme/communication/improvements-mid-2008/indicative-results/detailed-methodology-papers/index.html).

3c. How the estimates are calculated

This section details the methodology for the calculation of annual mid-year internal migration estimates and explores validation, assignment of geography, accumulation and duplicate records, identification of migrants, imputation, cross-border flows, constraining and migration of students.

Validation:

Once the PRDS data are received by ONS, validation checks are carried out to ensure that the data have been transferred to the ONS computer system correctly and to assess the quality of the data. Any records with a temporary NHS number are excluded as their details may not be reflective of their usual residence. Checks are also carried out to identify any records with incomplete data, for example missing NHS number or missing postcode. If a record fails one of the validation routines, it is defined as an incomplete record and goes through the imputation process discussed later.

Assignment of Geography:

The National Statistics Postcode Lookup Directory is used to match the postcodes on the record level data to local authorities and FHSAs. If a postcode is missing or cannot be found on the postcode file the record goes through the imputation process discussed later.

Accumulation and identification of duplicate records:

The data are combined and sorted by NHS number and acceptance dates. This allows the identification of duplicate records. These arise when a patient record is created in a new area before their details have been removed from a previous area. When this occurs, the record with the earlier acceptance date is removed, so that only one record is held on the patient registers. Other duplicate records may occur when a patient is unable to give sufficient information to the NHSCR to be identified and so a new NHS number is assigned to them. It is not possible to identify these duplicate records within the dataset.

Identification of migrants:

Following validation, the patient registers are combined to produce a total patient register for England and Wales. Migrants are then identified by comparing the current year's patient register with that of the previous year. This is done by matching on NHS number. By comparing the postcode from one year to the next, any patient that has a change of postcode is identified as a migrant.

The following example using fictional data shows how a migrant is identified:

Table 2: Identification of a migrant

Year	NHS No	Sex	DOB	Acceptance date	Postcode	Local authority	Health authority
2009	678456891	F	13041958	04051998	SE24 0AH	01AY-Lambeth	QAH-Lambeth, Southwark & Lewisham
2010	678456891	F	13041958	04112009	W1P 5RZ	01AG – Camden	QAT – Camden and Islington

The same person is identified in two consecutive years. In 2009, the postcode indicates that the person is resident in Lambeth with a postcode of SE24 0AH. In 2010, that same individual, with the same NHS number and same date of birth is resident in Camden with a postcode of W1P 5RZ. The move is also between Lambeth, Southwark and Lewisham and Camden and Islington former health authorities.

Some records will show a change of postcode that is simply a reclassification made by the Royal Mail. These changes of postcodes create false migrants and these are removed before the genuine migrants are identified.

Migrants are also identified where a patient is only on one patient register. Matches are made against data from the NHSCR, NRS and NISRA to identify cross border moves between the constituent countries of the UK.

Imputation:

When a record is incomplete, variables are imputed both during the validation process and after migrants have been identified using information from other similar records. Values for sex are imputed during the validation process alternately to males and females as missing values occur.

Age is calculated as 30th June of the current processing year minus the date of birth. Where the age for a migrant is invalid, it is imputed by matching on migrants with similar values in other variables. If other variables are also missing, age is imputed by matching on variables from migrants identified from incomplete records. If a complete match is not found, conditions are dropped until a match is made solely on the destination or origin FHSA.

If a postcode is missing, a migrant may still be identified if the FHSA of registration has changed from one year to the next. However, such migrants need to be allocated down to local authority level. An imputation method has been devised which is similar to that used for imputing missing areas of residence in the International Passenger Survey. The imputation is carried out by finding all other migrants with the same characteristics i.e. the same origin and destination FHSA, the same origin local authority and the same age and sex. Once these migrants are identified, the missing characteristics in the incomplete record is imputed using a random allocation of local authority from those migrants identified as being similar.

Using data for migrants between FHSAs where the postcode is known should not introduce any bias, since postcodes are not likely to be missing for migrants more than non-migrants.

If a postcode is missing and the FHSA of registration is unchanged from one year to the next, it is not possible to identify migrants. However, previous research indicates that the number of migrants within a health authority where a postcode may be missing or invalid is small. These migrants cannot be identified with the current methodology.

Cross-border flows:

Data from the patient registers only enables identification of people who have moved within England, within Wales and between England and Wales. However, the NHSCR, and data from NRS and NISRA can be used to provide information on moves from Scotland and Northern Ireland.

Migrants moving into England and Wales from Scotland and Northern Ireland are identified when the NHS number on the current year's patient register is not found on the previous year's patient register and the previous health authority on the current year's patient register indicates Scotland or Northern Ireland. These migrants are identified and then constrained to data from the NHSCR on moves into England and Wales from Scotland and Northern Ireland.

Migrants moving out of England and Wales to Scotland and Northern Ireland are identified initially where the NHS number on the previous year's patient register is not found on the current year's patient register. The NHS number is then matched to data from NRS and NISRA for persons moving from England and Wales to Scotland and Northern Ireland. If the NHS number is found on this data, a migrant to Scotland or Northern Ireland has been identified and the destination area is recorded as Scotland or Northern Ireland. These moves are then constrained to the data supplied by NRS and NISRA for moves into Scotland and Northern Ireland. This is because it is assumed that the country receiving the migrants is able to provide a more accurate count of the number of migrants.

Constraining patient register data to the NHSCR estimates:
The migration data based on the patient register are constrained to the NHSCR estimates at the FHSA level to account for migrants that may be missing.

The table below shows the effect of constraining on country level flows for years 2005-2009. It can be seen that while variation occurs in the constraining factor, the magnitude of constraining that occurs each year is similar. The smaller flows, such as Northern Ireland to Wales tend towards larger and more volatile constraining factors due to the small numeric nature of these flows and the large relative effect that small changes can have.

Table 3: Constraining factors* at country level; 2005-2009

IN COUNTRY	OUT COUNTRY	2005	2006	2007	2008	2009
England	England	1.16	1.16	1.15	1.15	1.19
England	Northern Ireland	1.11	1.11	1.13	1.12	1.17
England	Scotland	1.12	1.11	1.11	1.12	1.13
England	Wales	1.17	1.16	1.17	1.16	1.18
Northern Ireland	England	1.32	1.30	1.33	1.40	1.23
Northern Ireland	Wales	1.78	1.51	1.78	1.72	1.74
Scotland	England	1.38	1.25	1.27	1.35	1.27
Scotland	Wales	1.36	1.28	1.26	1.36	1.31
Wales	England	1.14	1.15	1.15	1.13	1.17
Wales	Northern Ireland	1.01	1.08	1.12	1.15	1.13
Wales	Scotland	1.11	1.11	1.10	1.12	1.14
Wales	Wales	1.13	1.13	1.13	1.13	1.17

*Base PRDS flows multiplied by this factor

Constraining migrants aged one and over:

The NHSCR holds information on moves of migrants aged one and over who have moved between FHSAs and this is used to constrain the patient register data for migrants aged one and over who move between FHSAs. Information on moves of migrants within health authorities is not available on the NHSCR data, which only records moves between FHSAs, and so it is assumed that the volume of scaling up moves between FHSAs is the same as that needed for within area moves.

For moves between FHSAs for migrants aged one and over, the number of moves for each combination of FHSAs in the patient register is constrained to the number of moves for the same combination of FHSAs in the NHSCR by applying scaling factors. To constrain the patient register data to NHSCR data for moves within a FHSA, the scaling factors used to constrain between area moves have been used. For each combination of FHSAs, the scaling factor for moves into that FHSA are combined with the scaling factor for moves out of that FHSA to obtain an average scaling factor for moves within that FHSA. This average scaling factor is then applied to all moves with each FHSA.This constraining process ensures that the resultant estimates includes the migrants who will not be on one of the two years patient registers.

Constraining migrants aged under one:

Comparing patient registers from two years will not capture any information on migrants aged under one since they will not appear on the patient register in the first year. While the NHSCR data does hold information on moves under one, this information is at FHSA level and the patient

register data is at a lower geographic level. However, research has shown that the distribution of male and female migrants aged under one by FHSA in the NHSCR data is very strongly correlated ($r2 = 0.97$) with the distribution of male and female migrants aged one by FHSA in the patient register data.

The distribution of migrants aged one in the patient register data by sex, origin FHSA and origin local authority is used to distribute data from the NHSCR on migrant babies under one by sex, origin FHSA and origin local authority level. Likewise, the distribution of migrant babies aged one in the patient register data by sex, destination FHSA and destination local authority is used to distribute data from the NHSCR on migrant babies under one by sex, destination FHSA and destination local authority level.

To constrain the patient register data for migrants aged under one within a FHSA, the ratios of within area moves to between area moves will be the same for migrants aged one and under one. Therefore to create migrants aged under one moving within a FHSA, the number of migrants aged one moving within a FHSA, which has already been constrained to NHSCR, is multiplied by the ratio of migrants aged under one to migrants aged one moving between areas.

Migration of students:

In 2010, ONS introduced an adjustment for students calculated from data from the Higher Education Statistics Agency (HESA). This adjustment was designed to better capture the movement of students to and from places of study.

The adjustment is applied to the constrained PRDS tables and consists of:

• an adjustment to the number of moves made by first year undergraduate students in the year to mid-2008 based on HESA data on students' domicile and term-time address

• an adjustment to the equivalent number of moves made by first year undergraduate students between mid-2001 and mid-2007 based on similar HESA data. Additional assumptions are made in the absence of HESA data on students' term-time address prior to academic year 2007/08

• an adjustment to the number of moves made by students at the end of their studies between mid-2001 and mid-2008. The approach uses a range of sources in the absence of HESA data on students' address in the year after their studies, and

• a double counting adjustment to avoid counting moves of students/former students twice, when they do eventually re-register

4. Reliability of Internal Migration estimates

Because there are no direct measures of migration in the intercensal period, proxy data are used. This means that any differences in coverage between the proxy and the phenomena that it is used to measure will affect the quality of the estimates. 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 both the migration estimates derived from the NHSCR and those made using patient registers directly.

This section covers the reliability of internal migration estimates and looks at the following; exclusion of some population groups, maintenance of the patient register, list inflation, accuracy of variables on the NHSCR, postcodes and variation in re-registration patterns.

4a. Exclusion of some population groups:
Neither data source can 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. It has been estimated as about 0.3 per cent of those sampled.

Certain groups of the population are excluded from the NHS such as Armed Forces personnel who receive treatment from Service Medical Officers. This includes the dependants of foreign armed forces. The NHS will also exclude prisoners that are sentenced a term of two years or more and patients in long stay medical hospitals when they have been in hospital for two years or more.

4b. Maintenance of the patient register:
In terms of maintaining the lists, the following rules apply, but the time taken to action these events will vary by PCT.

• When a person on a patient register dies, or is absent from the United Kingdom for a period of more than three months, they shall be removed from the doctor's list on the date on which the PCT first receives notification of the death or that the absence has exceeded three months.

- When a person on a patient register leaves the United Kingdom with the intention of being away for a period of at least three months; is in Her Majesty's forces; is serving a prison sentence of more than two years or sentences totalling in aggregate more than that period, they shall be removed from the doctor's list when the PCT first receives notification.

- When the PCT is satisfied that a person on the list of a doctor no longer resides in the locality, the PCT shall inform the person and the doctor that the doctor is no longer obliged to provide general medical services. After 30 days from the date of a letter, the HA will remove the person from the doctor's list.

- Where a person on the list of a doctor providing general medical services has moved to an address outside the locality, or the address of the person is no longer known to the PCT, the PCT shall inform the doctor that it intends at the end of 6 months to remove the person from the list and do so (this is known as the FP69 procedure).

4c. List inflation:

Patient registers from the individual health authorities form a comprehensive register of the population registered with NHS doctors, however research has concluded that they should not be used by ONS to estimate the resident population. The reasons for this are that some patients are registered in more than one area, that some patients may have more than one NHS number and finally that patients are on the doctors' lists after having died or left the country. The number of patients on the registers has exceeded the number of people resident in the country in every year since 1961. These problems do not affect the use of patient registers for measuring internal migration.

List inflation for each area can be calculated by comparing the population counts based on the patient registration data against the population estimates rolled forward from the 2001 Census. The difference between the two estimates, described here as list inflation is calculated using the following formula:

$$\text{List Inflation} = \frac{\text{Population count (patient registers) - Mid-year population estimate}}{\text{Mid-year population estimate}} * 100$$

The mid-year population estimates refer to the resident population, while patient registers give the population registered with a GP. Groups of the resident population that are not registered with an NHS GP and consequently excluded from the population counts are foreign armed forces and

their dependants, and home armed forces. Dependants of home armed forces should register with a NHS GP and therefore be included in the population counts. This difference in definitions of population will cause list deflation in areas where there are a significant number of armed forces.

4d. Accuracy of variables on the NHSCR:

NHSCR data are collected weekly, then processed and delivered quarterly on a rolling year basis. The variables held are the patient's NHS number, name, and date of birth along with the FHSA which the patient has or is registered with, along with acceptance dates for these FHSA. 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.

As part of the initial research to ascertain whether data from the patient registers could be used to estimate migration, the quality of the data was examined, in terms of completeness, coverage and accuracy. The research found that the quality of information held on the patient registers has improved steadily over time. By 2008-9 there were no records failing the range checks and the accuracy of variables on the patient register is thought to be very good.

There has been a significant improvement in the proportion of records failing validation range checks over time. For mid-2008 to mid-2009, the overall proportion of records failing these checks was less than 0.01 per cent.

4e. Postcodes on the patient registers:

The accuracy of postcodes on the patient registers is particularly important, since the definition of a migrant and the geography of a migrant's address is reliant on this field. Improvements in the quality of postcodes on the patient registers reflect the efforts by FHSA to improve the quality of the data on their registers. Between 1991 and 1999, the proportion of records with a missing postcode fell from 2.7 per cent to 0.4 per cent. The quality of data continued to improve over time and by 2009 the percentage of records with a missing postcode had dropped to less than 0.1% The proportion of records with an invalid postcode is virtually nil.

4f. Variation in re-registration patterns:

It is known that re-registration patterns vary by sex and age group. For example, young children, their mothers and the elderly usually re-register quite quickly after moving, while young men take longer to re-register than women of the same age. Women tend to use or expect to use the health

service more regularly and are therefore more likely to register with a GP when they move. Generally, young fit people are less likely to re-register with a GP when they move, and are most likely to be excluded from the patient registers.

As part of the initial assessment of the patient register data, the age and sex migration profiles were compared to the age and sex migration profiles identified from the NHSCR data and the 1991 Census. Comparison with the 1991 Census indicated that, for females, the levels and age patterns of migration between areas from the patient register data are very plausible. The same is true of the migration from the patient register data for males, apart from those aged between 16 and 29. The same findings were replicated when the 2001 Census was examined.

5. Future plans

The ONS is currently evaluating a new data source the Patient Demographic Service or PDS. The exact scope and format of any outputs will be determined after thorough testing of the new data source has taken place to verify its quality and stability.