Interim report on the potential use of Department for Work and Pensions data to improve population and migration statistics

May 2009

Summary

This is the first report on progress made to investigate the potential of Department for Work and Pensions administrative data to make improvements to population and migration statistics. Work will be subject to a continuous process of peer review, it will be considered by demographic experts and as methods are developed to make more use of administrative sources in estimation procedures, the results will be subject to review by reference panels of Local Authority experts. Key findings include:

- **Aggregate data from the Migrant Worker Scan will be one of the data sources included in the model being developed to improve the distribution of migrants to local areas. This model will be incorporated into the release of improved population estimates in May 2010.**

- **Comparisons show that there is good agreement between the pensions data and the mid-year population estimates. These data should be used in quality assuring the older age groups**

1. Introduction

1.1 ONS has now completed the first year of a five year cross government work programme to improve population and migration statistics. There are various strands to this work including one which is looking at the potential of alternative administrative data sources. We are carrying out feasibility studies with a number of government departments to consider how aggregate administrative data might be used with the aim of linking record level data once approval has been given for access. The data sources that we are investigating have the potential to make improvements to population and migration statistics, support the conduct of the Census in 2011 and contribute

to the Beyond 2011 initiative which will look at the data requirements of users beyond the 2011 Census and how they can be met.

1.2 Data sharing is essential for migration statistics improvement - this was identified in the Interdepartmental Task Force on Migration Statistics. However, sharing data between government departments requires formal agreement and sometimes legislation. The Task Force recommended investigation into a number of administrative sources including data on students from the Higher Education Statistics Authority, data on school pupils from the School Census and the Work and Pensions Longitudinal Study. Access to some of these sources needs to make use of the Statistics and Registration Act 2007. This was the case for data recently received from the School Census which we shall be reporting on in summer 2009. The data sharing clause in the Act allows secondary legislation (known as regulations) to be laid which enables data to be shared between government departments and the UK Statistics Authority. In some cases, we have found that legal gateways to data sharing already exist and we have not had to use the Act, as was the case for Migrant Worker Scan data.

1.3 ONS has jointly carried out a feasibility study with the Department for Work and Pensions (DWP) and working collaboratively, this has enabled expertise about specific DWP data sources to be shared across departments. An assessment of sources and aggregate outputs has been considered and the next stage is to consider how these sources can be used at record level either individually or through linkage with other sources. This paper reports on the progress of this work with DWP and makes a number of recommendations about how these sources can be used.

2. Potential data sources from the Department for Work and Pensions

The investigations into the potential of DWP sources identified a number of datasets which are considered below.

i) The Migrant Worker Scan

2.1 All workers are required to register with the Department for Work and Pensions for a National Insurance Number (NINo) and this includes all adult overseas nationals if they intend to work or claim benefits while living in the UK. These data are collected on the National Insurance Recording System (NIRS) which is owned by HM Revenue and Customs (HMRC). The Migrant Worker Scan is an extract of NIRS and holds information on all adult overseas nationals who register for a NINo in the UK. These data are provided to DWP for publication purposes using the data sharing provisions introduced in the Employment Act 2002.

2.2 The Migrant Worker Scan is widely thought to be a good source of information about migrants and that its potential should be investigated for making improvements to migration and population statistics. ONS initially analysed aggregate outputs from this source to better understand migrant

2

behaviour. This work identified the need for record level data so that further analysis and linkage with other sources could be explored. Working with HMRC and DWP, ONS has achieved access to record level data using an existing legal gateway (Section 122AA of the Social Security Administration Act 1992). Subsequently, ONS received extracts of the Migrant Worker Scan and work has begun to carry out record linkage. Three cumulative extracts covering all registrations up to June 2008 have been received so later extracts contain any newly registered migrants and take account of any new information received by HMRC e.g. changes in addresses or dates of death. Data has been provided back to 1975 but more recent data from 2002 onwards is considered to be of better quality.

2.3 Although the Migrant Worker Scan is a rich source of information, there are a number of coverage and definitional issues that need to be considered. These are described below.

- The Migrant Worker Scan does not hold any information on how long migrants stay in the country. It is possible that some migrants may arrive in the UK and work for a short period of time before returning overseas. This means that the data cannot provide a good stock count of migrants resident in the country. For instance, the data are not directly comparable with ONS estimates of long-term migrants i.e. those who stay in the UK for a period of 12 months or more or estimates of migration from the Annual Population Survey.

- The Migrant Worker Scan does not include all migrants. Groups that are excluded include child migrants, migrants who work illegally and adult migrants who neither work nor claim benefits. This last group are commonly spouses of employed migrants and students.

- Address information is provided when a migrant registers for a NINo at a Job Centre. Subsequent changes in address will only be updated on the National Insurance Recording System if HMRC are informed of the change. Anecdotal evidence suggests that HMRC are most likely to be aware of these changes for migrants who are claiming benefits or those who are self-employed and required to complete an annual tax return. Therefore, the address information for some migrants is likely to be out of date if they subsequently move following arrival in the UK and HMRC are not notified. In addition, initial quality assurance work carried out by ONS has indicated that a very small proportion of migrants may be providing a non-residential address when they initially register for a NINo e.g. contact details for a bank, employer or employment agency.

- When a migrant registers for a NINo, the registration is added to the Migrant Worker Scan. In some cases, a migrant may delay registering for a NINo for up to a year or more. This means that data from this source aggregated by year of arrival will be subject to revision each year as additional migrants already resident in the UK apply for a NINo.
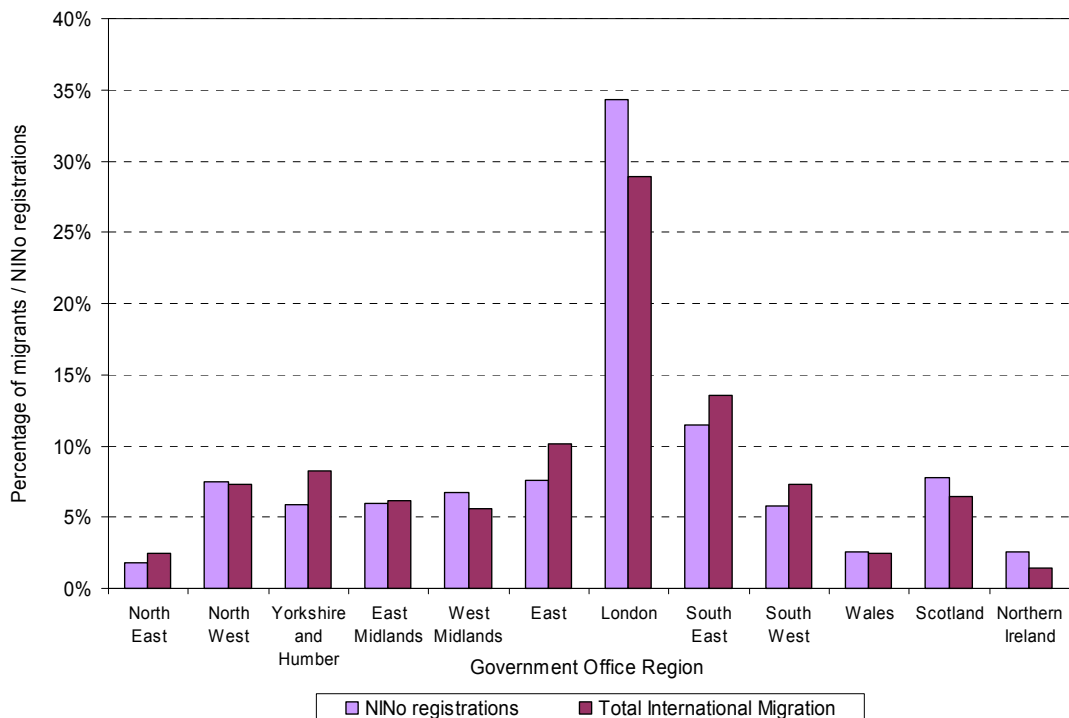
In addition, the date of arrival is self reported so this can introduce a further quality issue into data aggregated in this way.

Analysis of aggregate Migrant Worker Scan data

2.4 Aggregate data provided by DWP on National Insurance registrations from the Migrant Worker Scan by year of arrival have been compared to estimates of Total International Migration used in the mid-year population estimates. Data from the Migrant Worker Scan include short term migrants which are not included in the estimates of Total International Migration. These comparisons have highlighted both differences and similarities between the two sources.

2.5 There are variations in the age distribution of the two sources. Estimates of Total International Migration show a higher proportion of younger migrants aged under 25 and a lower proportion of migrants aged between 25 and 44 when compared to Migrant Worker Scan data. However, the distribution of migrants by sex is similar between the two sources. A comparison by Government Office Region is illustrated below and shows that in some areas, there are differences in the geographical distributions. For example, there is a higher proportion of NINo registrations in London than migrants estimated in the Total International Migration estimates. The differences that are shown here are likely to be due to definitional differences between the two sources.

Figure 1 – Distribution of NINo registrations by year of arrival compared to Total International Migration estimates by Government Office Region – 2006/07

2.6 These comparisons, together with further work, will help to develop our understanding of migrant workers and where they settle when they arrive in the UK. In addition, **aggregate Migrant Worker Scan data are one of the data sources included in the model being developed to improve the distribution of migrants to local areas.** This model will be incorporated into the release of improved population estimates in May 2010. Further work should include considering the implications for estimation at regional level and developing a better understanding of definitional issues.

ii) Lifetime Labour Market Database

2.7 The Lifetime Labour Market Database is a one per cent sample from the National Insurance Recording System which contains information on all persons making tax and National Insurance contributions. Migrants can be identified within this dataset as those persons who originate from abroad or enter the system later than their 16th birthday. This database has the potential to provide detailed information on the activity of migrants arriving in the UK using information on their tax and National Insurance contributions but only for a one per cent sample of migrants. It does not provide the level of coverage available from the Migrant Worker Scan, which includes all overseas nationals registering for a National Insurance number but does not include activity information on tax and National Insurance contributions.

2.8 The following coverage issues are associated with the Lifetime Labour Market Database.

- The Lifetime Labour Market Database does not include migrants who do not need a NINo, so child migrants, migrants who work illegally and adult migrants who don't work or claim benefits are not included. This last group are commonly spouses of employed migrants and students.

- The first tax year of activity together with the last date of any activity can provide a proxy for the period of time that the migrant has been resident in the UK. There are exceptions to this such as people who work in very low paid employment or the 'black market' economy, women who arrive, then marry and cease to work, people who retire or any other situation where tax and National Insurance are no longer paid.

- Address information on the Lifetime Labour Market Database is subject to the same quality issues as those on the Migrant Worker Scan since they both originate from the same source. The address is considered to be more up-to-date for migrants who are claiming benefits or those in self employment who are required to complete an annual tax return.

2.9 ONS needs to consider the Lifetime Labour Market Database further to decide if it can be used. Initial thinking is that it could help to inform short term migration estimates. We need to carry out further work to determine whether we would need to lay a regulation under the Statistics and Registration Act to gain access to these data.

iii) Work and Pensions Longitudinal Study

2.10 The Work and Pensions Longitudinal Study (WPLS) links data from HMRC on topics including employment, pensions and tax credits with benefit and back to work programme information from DWP. Data are collected for England and Wales with more limited information for Great Britain and the UK. This source is generally thought to have good potential to provide significant improvements to migration and population statistics since it covers a large proportion of the population.

2.11 ONS also recognise the potential of using the Work and Pensions Longitudinal Study at a higher level for migration and population estimation. It is considered to be a key dataset and ONS want to consider how it can be used for population and migration more generally and for the Beyond 2011 work which is considering data requirements of users beyond the 2011 Census.

2.12 The Work and Pensions Longitudinal Study is a complex collection of datasets and familiarity and legislative work is underway but it is recognised that this work will be complex and may take longer than for other sources such as the Migrant Worker Scan. DWP have advised that initial investigations should focus on aspects of the Work and Pensions Longitudinal Study rather than the dataset as a whole. The address information on the WPLS and data on pensioners have therefore been subject to further investigation for the reasons described below.

WPLS Addresses

2.13 The current address is held for all persons on the Work and Pensions Longitudinal Study. This information is kept up-to-date using an address hierarchy so that if more than one address exists for a client, a hierarchy of reliability is used to determine which address is likely to be correct. Self-employed and benefit clients are considered to have the most up-to-date address information.

2.14 Addresses on the WPLS have been identified as having two uses. Firstly, as an additional list of addresses for use in the 2011 Census, either in the enumeration, validation or quality assurance process. Secondly, WPLS addresses may be able to provide information on internal migration. Further work is underway to determine whether ONS would need to lay a regulation under the Statistics and Registration Act to gain access to these data.

Data on pension claimants

2.15 The Work and Pensions Longitudinal Study holds information on all persons claiming either a state pension or pension credit. Some women aged 60 to 65 are not included because they have no pension entitlement of their own and will only receive a pension when their husbands turn 65. Other small groups of the population will also be excluded from this source if they have not paid sufficient National Insurance contributions. This includes international migrants who have retired but not paid sufficient National Insurance contributions to be eligible for a pension. Pension claimants who have moved abroad and notified DWP can also be identified.
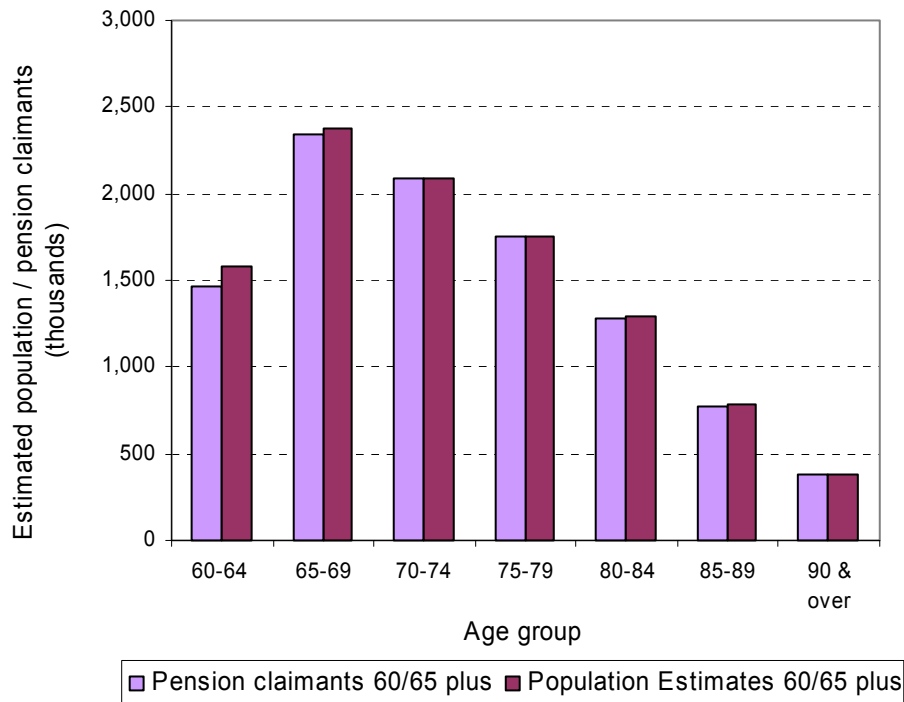
Analysis of aggregate pensions data

2.16 DWP have provided aggregate data on all males aged 65 and over and all females aged 60 and over claiming a benefit. The data are referred to as data on pension claimants since most of these claims will be for a state pension. However, there will be claims for other types of benefit including pension credit and attendance allowance. Comparisons between pensions data and estimates of the population of pensionable age have highlighted similarities between the sources.

2.17 The total number of pension claimants compared to the population estimates for men aged 65 and over and women aged 60 and over is shown in Figure 2. The total number of pension claimants expressed as a percentage of the mid-year population estimates is 98% and when women aged 60 to 65 are excluded from this analysis, a group known to suffer from undercoverage in the pensions data, this figure increases to 99%. Aggregate analysis has found that the coverage is broadly consistent across Government Office Regions although there are a small number of Local Authorities (2% of areas) where the number of pension claimants expressed as a percentage of the mid-year population estimates falls below 90%. In two of these areas (both London Boroughs), this proportion falls below 80%. This analysis needs further investigation which may require access to record level data.

2.18 **Overall, these comparisons show that there is good agreement between the pensions data and the mid-year population estimates. This data source should be used in quality assuring the older age groups** and potentially could be used directly in the estimation process although we would need to gain a better understanding of local distributions.

Figure 2 – Pension claimants compared to mid-year population estimates – 2007

3. Next Steps

3.1 The work described above has meant that ONS and DWP have been able to share expertise about DWP administrative sources and has enabled ONS to identify how at aggregate level they can be used to make improvements to population and migration statistics. It is important that we have a good understanding of these sources so that we can produce a strong statistical business case, which is essential for access to record level data. We are now looking at obtaining access to more record level data and determining whether regulations are needed.

3.2 There are a number of ways in which DWP administrative sources could be of potential use in population and migration estimation. The sources could help us to determine whether we are getting the population estimates right between each Census for those groups that are difficult to estimate e.g. the younger age groups. We might then be able to use the Work and Pensions Longitudinal Study to make an adjustment for those population groups that are difficult to estimate. We would also like to compare migrant flows captured in the Migrant Worker Scan with other sources of information on migrants which might lead to an adjustment to our international migration estimates.

3.3 To establish how we can use DWP administrative sources for population and migration estimation, we intend to carry out some exploratory work. The next steps therefore involve several strands of work which we have either begun or we have planned for the future. Work which we are doing or will begin shortly includes:

- Further investigation of record level Migrant Worker Scan data to improve our understanding of migrants and their behaviour.

- Further comparisons of Migrant Worker Scan data with migration estimates e.g. combined estimates of long and short-term migration for more meaningful comparisons.

- Linking subsequent extracts from the Migrant Worker Scan to capture onward migrant moves. While many migrants will not inform HMRC of address changes, it is hoped that sufficient changes will be captured to better understand distributions of onward moves.

- Testing the feasibility of matching Migrant Worker Scan data to patient register data which could show whether migrants are still resident in the country and where they are living. We do not expect this matching to be straightforward because the unique identifiers on each source are different.

- Production of a feasibility report on the usefulness of Migrant Worker Scan data, as agreed in the Memorandum of Understanding between ONS, DWP and HMRC.

- Considering whether record level data from the Lifetime Labour Market Database should be pursued.

- Determining what sources we might obtain from the Work and Pensions Longitudinal Study. A workshop between ONS, DWP and HMRC has been arranged to discuss this source and determine whether record level data are required. Preparatory work is already underway to establish whether a regulation is required, should we need access to record level data. Potentially, WPLS addresses could inform internal migration estimates and data on pension claimants could improve the internal migration estimates of the population of pensionable age.

- Exploring the potential use of record and aggregate level data to quality assure Census estimates. This includes assessing the potential of data from the Migrant Worker Scan, Lifetime Labour Market Database and Work and Pensions Longitudinal Study for Census Quality Assurance. Wherever possible, analyses already undertaken within ONS will be utilised to avoid duplication of effort. DWP data sources may be used in the following areas for Census Quality Assurance in 2011:

  i) At aggregate level to compare with Census estimates.
  ii) To assess the coverage and completeness of DWP data sources to evaluate their suitability and inform their use as a Census comparator at aggregate level.

iii) Along with a range of other administrative sources, data linkage of DWP data sources are being considered as a means of filling information gaps identified through Census validation.

- The Beyond 2011 programme will make an assessment of the DWP Customer Information System to explore the possibility of contributing to the creation of a 'spine' of the population, in combination with other data sources such as WPLS addresses or other administrative data sources. This would involve record-level linkage of data, and other similar work in ONS will be taken into account.

Work which we plan if we gain access to additional record level data includes:

- Linking Migrant Worker Scan data with other administrative sources to explore whether these sources, when considered in combination, provide good evidence of migrant activity. For example, linkage to tax and National Insurance activity on the Work and Pensions Longitudinal Study could provide information on migrants' length of stay.

- Linking Migrant Worker Scan data to the Lifetime Labour Market Database if we decide this is helpful. This would provide additional activity information on up to one per cent of the Migrant Worker Scan records depending on the success of the linkage and provide an estimate of length of stay, albeit not for the full dataset.

3.4 As the improvement work to identify the potential of alternative administrative sources continues, further papers will be published on the Improving Migration and Population Statistics webpage (http://www.statistics.gov.uk/about/data/methodology/specific/population/future/imps/updates/default.asp) to inform users of progress.

Annex A - Summary of DWP data sources

Source	Description	Coverage	Potential Uses	Comments
Migrant Worker Scan (MWS)	An extract of the National Insurance Registration System of all adult overseas nationals registering for a National Insurance Number.	All adult overseas nationals who wish to work or claim benefits while resident in the UK.	Estimate of inflow of short and long-term migrants combined, by geography, age and sex. Linkage with other sources to improve our understanding of migrant activity including onward moves and length of stay. Aggregate data has been included as one of the data sources in the model being developed to improve the distribution of migrants to local areas.	No information on how long a migrant stays in the UK. Excludes children, students and spouses who do not work or claim benefits and migrants working illegally. Addresses are only updated if HMRC are informed. Anecdotal evidence suggests that migrants who are self employed or claiming benefits are more likely to inform HMRC. Some migrants may delay registering for a National Insurance number. Therefore, data aggregated by year of arrival will be subject to revision.
Lifetime Labour Market Database	A one per cent sample of the National Insurance Recording System (NIRS).	Contains information on tax and National Insurance contributions on a one per cent sample of persons with a National Insurance number resident in the UK.	Linkage with other sources to improve understanding of migrant activity including onward moves and length of stay.	Migrants can be identified as those who originate from abroad or enter the NIRS system later than their 16th birthday. Does not provide complete coverage of migrants because it is a one per cent sample. Excludes those migrants who do not need a National Insurance number (also excluded from the MWS). Tax and NI activity can provide a proxy for residency. However, there are exceptions to this e.g. those in low paid employment, women who marry and cease to work, the retired etc. Address information is subject to the same quality issues as address information on the MWS.

Summary of DWP data sources - continued

Source	Description	Coverage	Potential Uses	Comments
Work and Pensions Longitudinal Study (WPLS)	A longitudinal dataset linking HMRC data on topics including employment, pensions and tax credits with benefit and back to work programme information from DWP.	Data collected for England and Wales. More limited information available for the UK.	Estimate of population by age, sex and geography. Changes in data could give components of population change.	The WPLS is a complex collection of datasets. Familiarity and legislative work is underway but may take longer than for other sources.
Address information from the WPLS	An extract of address information for all persons on the WPLS.	Data available for England and Wales. More limited information available for the UK. Address data held by DWP for DWP customers is considered to be more comprehensive.	To inform internal migration estimates. As an additional list of addresses for use in the 2011 Census, either in the enumeration, validation or quality assurance process. Linkage with other sources to produce a population 'spine' for the Beyond 2011 programme.	Addresses are kept up-to-date using an address hierarchy. Address information is considered to be most up-to-date for the self employed and benefit clients.
Pensions data from the WPLS	An extract of all DWP benefit claimants aged 60/65 plus from the WPLS.	All women aged 60 plus and all men aged 65 plus resident in GB and in receipt of a state pension or other DWP benefit.	Estimate of population of pensionable age by age, sex and geography. To quality assure current population estimates of pensionable age. Inform migration estimates of the population of pensionable age.	Undercoverage of women aged 60 to 65 due to some women having no pension entitlement of their own. Persons who have paid insufficient National Insurance contributions will not be eligible for a state pension e.g. some migrants. Includes pension claimants who have moved abroad but failed to notify DWP. DWP are not always notified of deaths which occur abroad. Data for the older age groups are therefore considered less reliable.