

# Improved Immigration Estimates to Local Authorities in England and Wales: Overview of Methodology

## 1. Introduction

This report provides an overview of an improved methodology for estimating long-term immigration to local authorities (LAs) in England and Wales. The new approach is based on using administrative data sources to distribute the England and Wales immigration totals from the International Passenger Survey (IPS) directly to LAs. The approach splits the IPS into different streams, mainly by 'reason for migration' (e.g. worker, student, other) and then maps each stream to the most relevant administrative sources which are then used to distribute immigrants to each local authority. For example, workers are distributed using National Insurance (NINo) data from the Department of Work and Pensions (DWP); students are mainly distributed using Higher Education Statistics Agency (HESA) data, while children and some other migrants are distributed using 'Flag 4s'<sup>1</sup> from the GP patient register data (PRD).

The improved methodology has been developed as part of the cross-government Migration Statistics Improvement Programme (MSIP) and replaces the 'model-based' approach that was developed earlier in the Programme and implemented in May 2010. The report provides some background to, and an overview of, this new methodology. There are four supplementary reports that provide additional detail on how each of the four main 'streams' (i.e. Workers, Students, Returning Migrants and Other) are distributed. Links to these reports can be found under the '*Detailed Long-term Migration Methodology Papers*' heading on the main release page.

This report starts with a review of the current methodology (Section 2) and the background to the development of a new approach (Section 3). This is followed by an overview of the improved methodology (Section 4) and each of the major components of the methodology (Sections 5-8). The report concludes by presenting the evidence as to why this method is an improvement over the current methodology (Section 9).

---

<sup>1</sup> A Flag 4 is assigned to those who register with a GP and whose previous address was outside of England and Wales

## 2. Overview of the current methodology for estimating immigration to local authorities

### 2.1 Distributing from a national immigration total

ONS estimates of international migration are based on the International Passenger Survey (IPS). This is a long running ONS survey that operates at UK ports of arrival and departure. The IPS is the only source on UK migration that is specifically designed to identify people who meet the UN definition of an international migrant, that is, someone who changes their country of usual residence for at least 12 months. This is also consistent with the 'usual residence' definition that is the basis for official population estimates.

However, there are two main weaknesses with the IPS. The first is that immigration estimates are based on a relatively small sample (approximately 2500 a year), and so the estimated flows cannot be disaggregated directly to LA level. For this reason IPS migration estimates are 'distributed' to lower level geographies using other sources and methods.

The second main weakness is that the migrant data in the IPS is 'intentions based', which means that a migrant's initial intentions about where they will settle may not be realised. Comparisons between 2001 Census data and the IPS suggest that there is a tendency for migrants to state an intention to settle in London, but actually settle in another part of the UK. Similarly, within regions there is a tendency for migrants in the IPS to state an intention to settle in regional centres; whereas the Census suggested that some of these migrants settle in nearby, less-well known LAs. This bias effect is commonly referred to as 'centralising tendency'.

### 2.2 Immigration distribution method used between 2007 and 2010:

In 2007, a number of improvements were made to the methodology for estimating immigration, which attempted to address these issues of centralising tendency (ONS, 2007). The key elements were:

- The replacement of an obsolete health geography with a new 'intermediate geography' called the 'new migrant geography for in-migrants' (NMGi) based on groups of LAs within regions. These were designed to deal with issues of centralising tendency to well known regional cities, but also to be large enough to produce a statistically robust IPS total that could be distributed to lower geographies.
- Constraining LA immigration estimates to Labour Force Survey (LFS) regional migrant distributions in order to better represent the regional distribution of where migrants settle therefore addressing the issue of centralising tendency at the national level.
- Use of 2001 Census migration data to distribute immigration flows down to LA level.

The method for distributing outside of London is shown in Figure 2.1. The method for distributing within London is slightly different and is shown in Figure 2.2. All of these elements, apart from the use of Census data, are part of the current method. In 2010, the 2001 Census distributions were replaced by a modelling approach incorporating more recent data sources.

Figure 2.1  
**Previous method for distributing immigration to LA level outside of London**

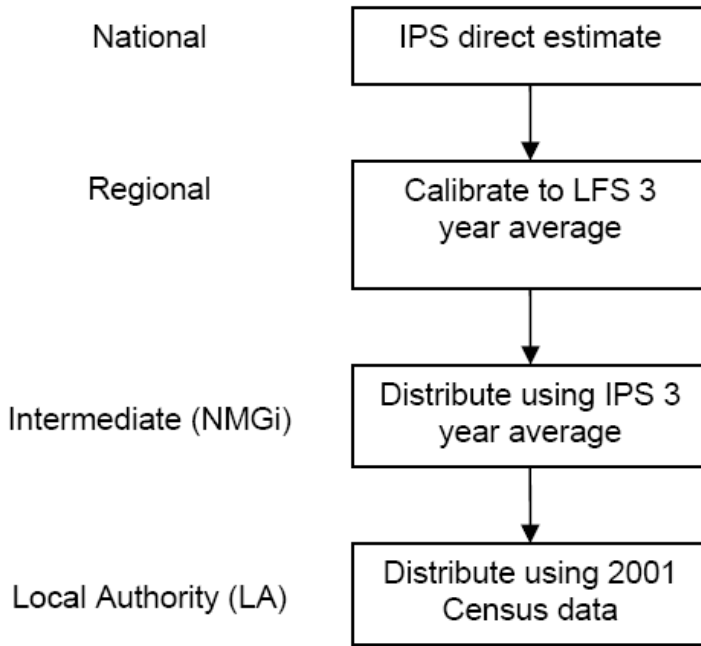
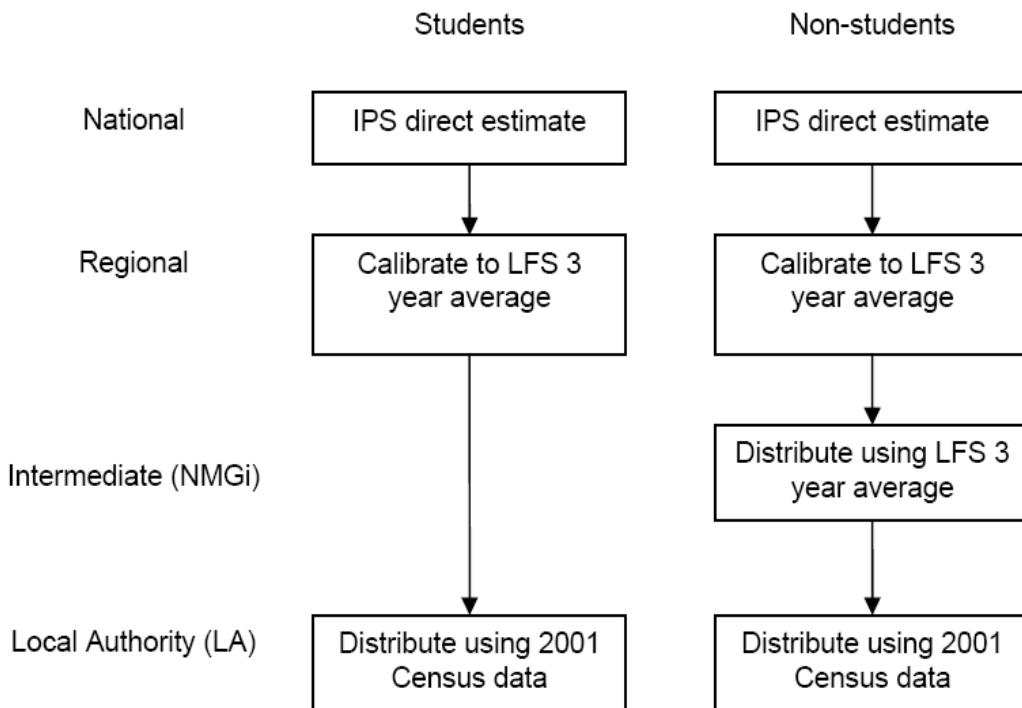


Figure 2.2  
**Previous method for distributing immigration to LA level within London**



**Note:** The method described above using the 2001 Census data was replaced by a modelling approach in 2010. This is explained further in section 2.3.

## 2.3 Phase 1 Improvements to the method for distributing immigration (introduced in 2010)

Patterns of immigration and settlement have changed considerably since the 2001 Census. For example, the large increase in A8 migration (i.e. from the eight Central and Eastern European countries that acceded to the EU in 2004) has seen increased immigration to areas that did not historically receive large numbers of migrants. A weakness of the methodology used between 2007 and 2010 is that use of 2001 Census data was not able to capture these trends.

In May 2010, this issue was addressed by implementing a new 'model-based' approach. This approach produces estimates for domains (in this case LAs) for which survey data is insufficient (due to small sample sizes) by borrowing strength from other data sources (ONS, 2009a). The other data sources, known as covariates, are available for all local areas. The covariate data is taken primarily from administrative data sources, but also includes some data from the 2001 Census, survey sources and modelled estimates.

The local authority estimate is based on the area-level relationship between the survey variable and the covariates. This relationship can be fitted by regressing the survey responses of the number of immigrants going to a LA, on area-level values of the covariates, e.g. number of new Flag 4 counts. The fitted model describes the relationship between the area-level survey variable and the covariates and this is assumed to apply nationally. Since these covariates are known for all LAs, the fitted model can then be used to obtain a full set of estimates. These modelled estimates will be more precise than the direct survey estimates<sup>2</sup>.

As part of the development process, the methodology was presented to an Academic Reference panel and the estimates were sense checked by a Local Authority Reference Panel (LIRPS). The Academic Reference panel raised concerns around the complexity of the modelling approach and the logic of constraining model based estimates to NMGis. There were also concerns around the regional estimates and the use of the LFS. The LIRPs broadly acknowledged these as being an improvement over the Census based estimates. They gave higher estimates for LAs that were known to have high levels of A8 migration (e.g. Peterborough, Boston). However, the method continued to produce estimates for some LAs (e.g. Newham and Brent) that did not address inconsistencies between the modelled estimates and administrative sources.

Given that the method was generally considered to be an improvement, ONS implemented it by May 2010 so the modelled estimates could feed into the 2008-based sub-national population projections. It was agreed that the overall method for distributing immigrants would be revisited in the next phase of the MSIP to address the remaining inconsistencies that existed between the modelled estimates and the administrative sources.

## 3. Distributional approaches using administrative data

### 3.1 Early research into the potential of administrative data

Before the MSIP launch, ONS had undertaken research into the potential of administrative data sources to improve the estimation of immigration focusing on NINos, Flag 4s and Worker Registration Scheme (WRS) (ONS, 2007). This work confirmed many of the issues already known to exist with administrative sources such as differences in coverage and inconsistencies with the UN definition of long-term migration. There was also a concern that the WRS was not expected to be maintained in the longer term.

---

<sup>2</sup> Further information on the model-based approach that was developed is available from:  
<http://www.ons.gov.uk/ons/guide-method/method-quality/imps/msi-programme/communication/improvements-mid-2008/methodology-papers/immigration-detailed-methodology-update.pdf>

The conclusion was that, without properly understanding the relationship between ONS definitions and administrative sources, it was not possible to identify which source produced the best results. This underscored the importance of gaining better access to administrative data, including micro-data, and undertaking the necessary research to determine how best to use them.

In 2008, during the early part of the MSIP there was some further consideration of administrative data as a possible alternative approach to the replacement of the 2001 Census distributions (ONS 2009a). Attention focused on Flag 4 and NINo data sources looking at both complete and Partial Replacement<sup>3</sup> of the 2001 Census data and both direct and ratio change methods. Various approaches were taken to evaluate the results of the research including comparison of:

- 2001 estimates for direct methods with 2001 Census
- 2002 estimates for all methods with 2001 Census, and
- Local authority estimates aggregated to NMGi with existing NMGi totals for 2001-2005

The various methods were applied to a selection of local authorities known to have high levels of immigration. The results showed that the methods that had produced the most plausible results varied from area to area, so that no single approach performed better than the others. This was thought to be partly due to differences in the characteristics of long-term migrants across the country and how well they were captured by the respective administrative data sources.

It was determined that the best way of making use of the administrative sources was to make them available for selection as covariates in a model based approach. In this way, the strengths of the administrative data could be used without making subjective judgements about how they should be combined. This was the basis of the model-based methodology implemented in May 2010.

### 3.2. University of Leeds approach

Around the same time that the modelling approach was being developed, an alternative method using administrative data was also being developed by the University of Leeds as part of the development of the Economic and Social Research Council (ESRC) funded 'New Migrant Databank'. The approach involved distributing the Long-term Total International Migration (LTIM) figure (i.e. the IPS estimates plus Irish flows, asylum seekers and 'switchers'<sup>4</sup>) for England and Wales to local authorities using different administrative sources (Boden and Rees, 2010).

A number of models were developed and two were presented in detail. 'Model A' was based entirely on Flag 4 data while 'Model B' was more complex and used a combination of Flag 4, NINo allocations to foreign nationals, and data from the HESA. The main innovation of the latter was to use the IPS reason for migration data (i.e. work, students, and others) to identify separate flows, which are then distributed using the administrative sources that correspond to those migrant flows. This approach largely addressed the problem that administrative sources often only cover a subset of the population. The main limitation of this approach was that it used a total flow based on a long-term definition and distributed it using sources that did not correspond to the UN definition of a long-term migrant. For example, the data being used contained both short and long-term migrants.

---

<sup>3</sup> Partial Replacement describes the use of an administrative source which is restricted to the migrant population covered by that source (e.g. Flag 4s to the non-UK born population)

<sup>4</sup> "Visitor switchers" are those who arrived intending to stay for less than 12 months but actually stayed longer, "migrant switchers" are those who arrived intending to stay 12 months or longer, but actually stay less

### 3.3 An independent review of methods

The direct distributional approach developed by the University of Leeds sparked a debate about whether it was better than the modelling approach proposed by ONS. To address these concerns ONS commissioned an independent review of methods for distributing international immigration estimates. The report summarised the debate as:

*“... a trade-off between the adherence to common definitions with respect to the duration-of-stay criteria but at a risk of lower coverage of migrant populations [in the ONS method], and more complete and up-to-date coverage, at the expense of some definitional differences [(in the New Migrant Databank)].” (Bijak, 2010: p.18)*

The review recommended that ONS seek a compromise and make greater use of administrative sources where possible. Suggestions included future comparisons of distributions based on administrative sources with 2011 Census, and consideration of approaches that combined both survey and administrative data.

The MSIP Board agreed that immigration methodology should be revisited in the next phase of the Programme.

### 3.4 Building on a direct distributional approach using administrative data

Considerable progress was made in gaining access to administrative record level data during the first phase of the Programme and understanding the potential of these sources. These included data from HESA, the School Census and the Migrant Worker Scan<sup>5</sup> (MWS). In addition, in the first half of 2010, arrangements were put in place to give ONS access to the Lifetime Labour Market Database (more commonly referred to as the ‘L2’) held by DWP. This data source provided the opportunity to identify activity associated with a NINo, which could in turn enable migrant workers to be more precisely mapped to the ONS definition of a long-term migrant. This would help address the main limitation of the University of Leeds distributional approach, namely, that published data on administrative sources do not correspond to the ONS definition of a long-term migrant. Following a brief feasibility assessment, it was determined that it was possible to develop a methodology that would be an improvement over the existing approach. The following section gives an overview of the methodology that has been developed.

---

<sup>5</sup> The MWS is a record of those from outside of the UK who have registered for a NINo.

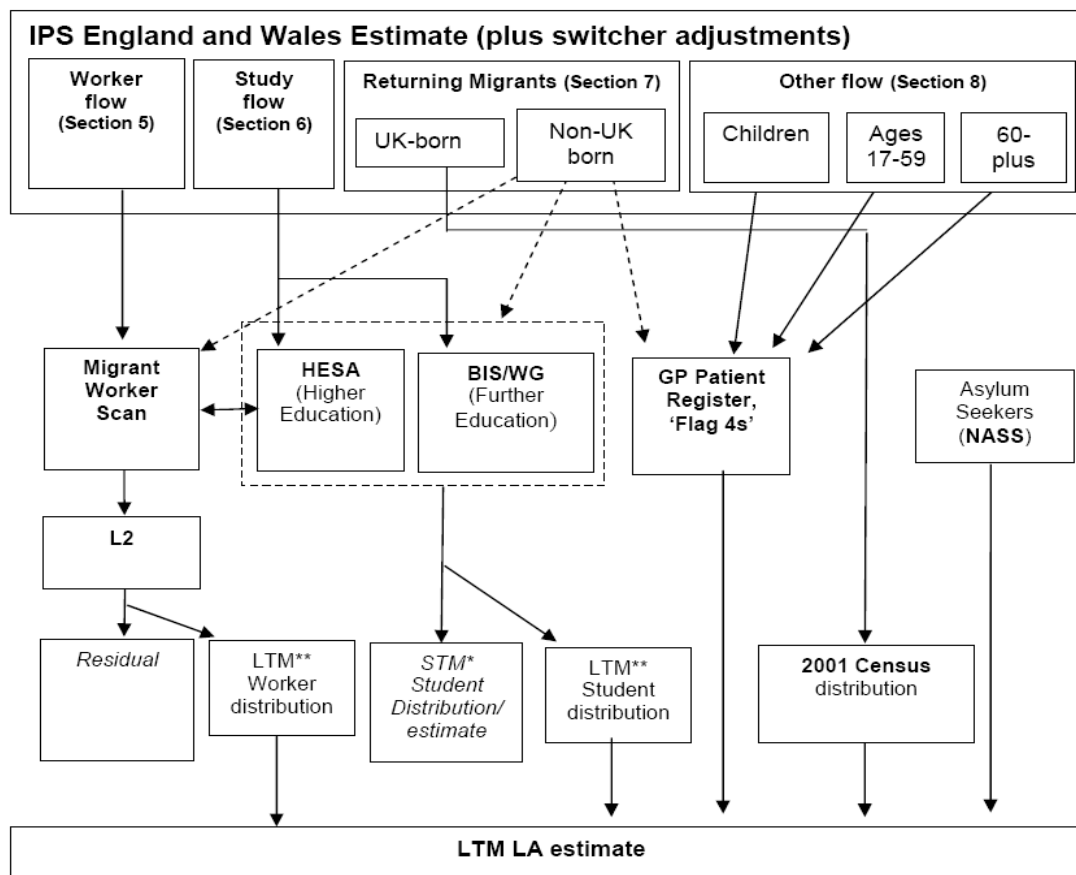
## 4. The Improved Immigration Distributional Methodology

This section provides a high level description of the improved methodology. The subsequent sections provide further details for each of the major elements of the Methodology.

### 4.1 Overview of the Improved Distribution Model

An overview of the distributional model is presented in Figure 4.1:

Figure 4.1  
Overview of New Immigration Distributional Model



\*STM = Short-term migrant  
\*\*LTM = Long-term migrant

Note: A Flag 4 is assigned to those migrants who register with a GP and whose previous address was outside of England and Wales

The main features of the new distributional methodology are as follows:

- The key principle is to achieve the closest possible mapping between the IPS and the available administrative data
- The LA estimates are based on distributions and not the actual administrative counts. Thus, the total population estimate for England and Wales does not change<sup>6</sup>
- A distinction is drawn between 'first-time' migrants and 'returning' migrants because of differences in the way in which they interact with the administrative sources

<sup>6</sup> This is the intention for the indicatives to be produced this year. In future years, the method will incorporate Scotland but this needs to be worked through with National Records of Scotland (NRS).

- Record linkage is used both within and between sources to minimise definitional differences and double counting. This is covered further under Section 4.3

The sources used are as follows:

- **Migrant Worker Scan (MWS)** provides a count of foreign nationals applying for a NINo
- **Lifetime Labour Market database (L2)** is used to estimate the proportion of the NINo count who are long-term migrant workers
- **HESA administrative data** is used for distributing publicly funded Higher Education student flows
- **HESA survey data** is used to distribute private Higher Education flows
- **Department of Business, Innovation and Skills (BIS) Welsh Government (WG)<sup>7</sup>** are administrative data sources used to distribute Further Education student flows
- **2001 Census data** for distributing UK-born returning migrant flows.
- **National Asylum Support Service (NASS)** data to distribute asylum seeker flows identified in the IPS<sup>8</sup>; and
- **Flag 4** data from the GP Patient Register Database, to distribute the remaining migrants

The intention is that the method will be applied for mid-2005/06 to mid-2009/10. This is partly because this five year period will contain sufficient trend data for the 2010-based SNPPs (Sub-National Population Projections), but also because some of the administrative data are not available for earlier years.

## 4.2 Key Assumptions

There are three fundamental assumptions underpinning the improved methodology:

*Assumption 1: The IPS immigration estimate is the best possible estimate at the national level*

The justification for this assumption is that the IPS is the only source that is specifically designed the measure migration based on the UN definition of an international migrant and is the basis for official estimates of international migration.

The administrative data is only used to *distribute* IPS flows down to regional and LA level. The improved approach seeks to map IPS flows as closely as possible to the most appropriate administrative source, and so it is the combination of the IPS flows and the distribution across the sources that will determine the LA immigration estimates, not the administrative counts themselves. This means that there will inevitably be differences between migrant counts from the administrative sources and corresponding IPS flows.

*Assumption 2: The 'main reason for migration' data in the IPS is a suitable basis for categorising the national IPS estimates so that they can be distributed down to LA level.*

The methodology hinges on being able to map the IPS to the administrative data. However, it is known that individual intentions behind decisions to migrate are complex and that the relationship between the IPS and the administrative data may not always hold in practice. For example:

- Someone whose main reason for immigration is to accompany or join their spouse could enter paid employment after arriving in the UK, and thus appear as a working migrant in the Migrant Worker Scan (MWS) and L2

<sup>7</sup> The data from BIS and WG is essentially the same, but is administered by the respective agencies for England and Wales.

<sup>8</sup> The methodology for distributing asylum seekers is unchanged but is included for completeness



- A migrant coming to the UK stating an intention to study, could enter full-time employment and never actually enrol in their course
- An EU citizen could come to the UK intending to work, but find out that they do not have the relevant qualifications and then decides to enter Higher Education whilst working part-time
- A migrant could arrive with the intention of working, registers with NINo but is then unable to find work.

Although the stated reason for migration will not hold in all cases, the reason for migration flows in the IPS can generally be reconciled with the administrative counts in the administrative sources<sup>9</sup>. There are some groups for which the gap between the IPS and the administrative data is difficult to explain, the most important examples being that the IPS identifies considerably fewer child migrants than those indicated by the Flag 4 data as well as fewer long-term A8 migrants than suggested by the L2 and MWS. However, with respect to A8 migration, the IPS and the administrative sources are consistent in that both indicate that the vast majority of A8 migrants are workers. For children, all are distributed using Flag 4 data.

For most IPS 'reason for migration' flows, the corresponding administrative data is broadly plausible and, where it is difficult to explain, the flows are either mostly or completely of the same type (and therefore distributed using the same source). Therefore, the IPS 'reason for migration' data is considered to be the sound basis for mapping to administrative source in order to distribute the England and Wales IPS total to LAs.

*Assumption 3: Any differences between the definitions used to map the administrative sources and the corresponding IPS data do not introduce geographic bias.*

Although the improved methodology has been able to refine some of the data sources to better align with the UN definition of a long-term migrant, it is not possible to achieve this with precision. If this were possible, then the administrative counts would be used directly rather than being used to distribute the IPS data. However, any mismatches between the administrative data and the UN definition will only affect the estimates if there is a geographical bias within these differences. Although it is possible some unknown bias may exist, these are likely to be much smaller than the known biases of the current method<sup>10</sup>. Therefore, as a working assumption it is assumed that any differences between the definitions used to map the administrative sources and the corresponding IPS data will not introduce geographic bias.

### 4.3 Record Linkage

The improved methodology incorporates some record linkage with the principle aim of reducing double counting both within and between sources. For example, international students who work could be captured in HESA, the MWS as well as the PRD. These data sets can be matched using date of birth, sex and postcode to identify administrative records that are likely to be the same individual. Similarly, there is some longitudinal linking both across and within data sets. For example, linked HESA data across years based on the unique student identifier has been used to identify international students arriving at a university, but who have a recent history of studying in the UK and therefore would have been counted as an immigrant in an earlier reference period.

Sensitivity analysis has showed that such instances of double counting can materially affect the local authority immigration estimates. In such cases linked records are removed from data sets so that, as far as possible, a linked record will only be counted once within the appropriate reference year. This approach will thus minimise any bias associated with double counting.

---

<sup>9</sup> Evidence to support this is presented in the sections detailing the methodology for each major work stream.

<sup>10</sup> These are discussed further in Section 9

It is important to recognise that this record linking done as part of the methodology is not comprehensive. The use of date of birth, sex and postcode as the match keys mean it is not possible to link sources where a migrant changes address between registering with one source and another. However, the linking approaches developed can be shown to deliver improved distributions, and as such, is an important element of the overall methodology.

#### **4.4 Method for deriving the IPS totals to be distributed**

There are a number of steps used to split out the IPS totals that are then distributed by the various administrative sources. These are designed to::

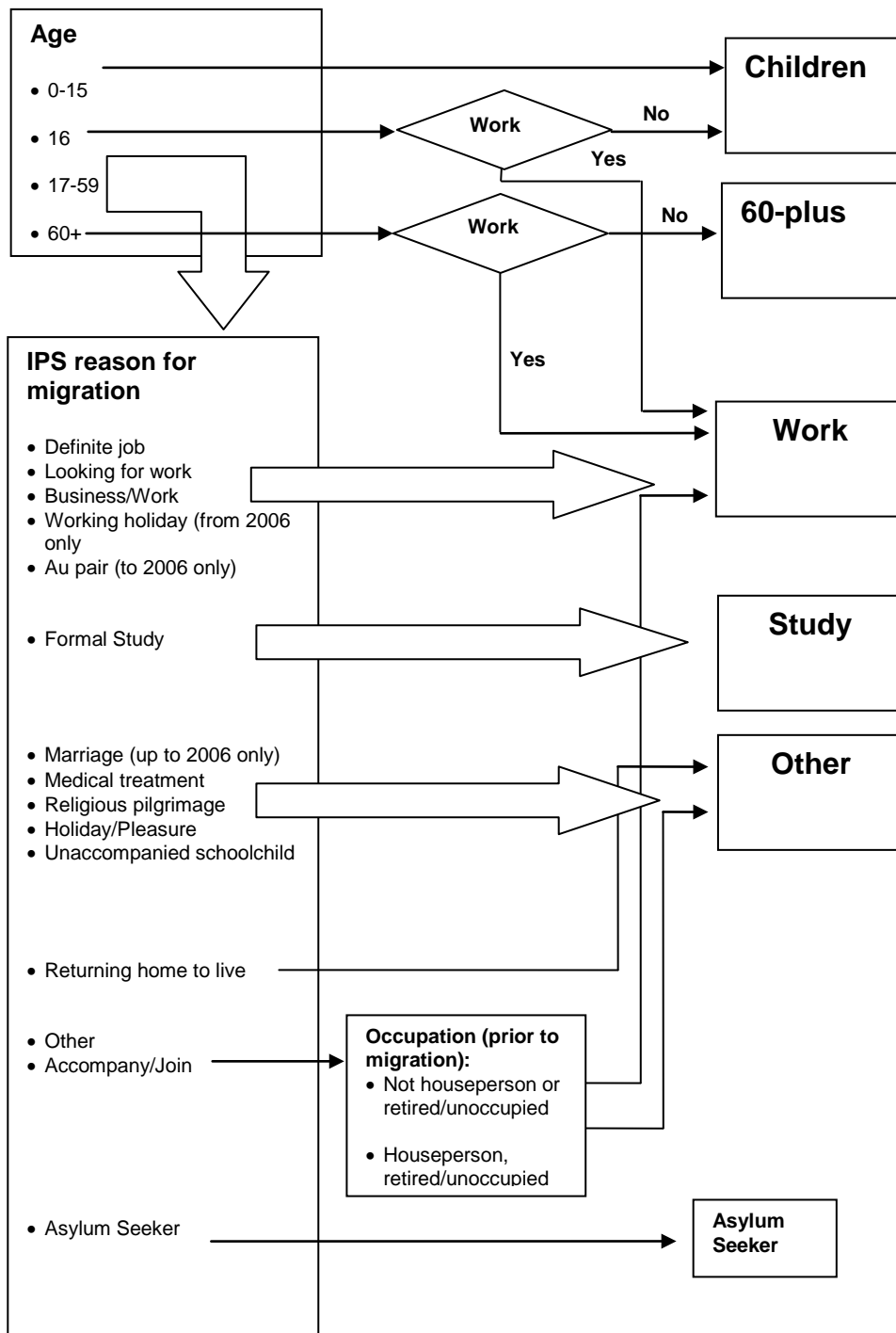
- achieve as closely as possible the mapping of the IPS flows to the administrative data sources.
- ensure that the other elements of LTIM (i.e. including migrant and visitor switchers, flows to and from the Republic of Ireland and Asylum Seekers are incorporated appropriately)
- ensure that the total England and Wales<sup>11</sup> flow is consistent with published estimates

The IPS asks respondents about their 'main reason' for coming to the UK. Figure 4.2 illustrates how first time-migrants to the UK, returning long-term migrants and the visitor switchers are grouped into 'work', 'study' and 'other' based on reason for visit. The methodology for preparing the IPS to be distributed is presented in Annex A.

---

<sup>11</sup> This is currently done in order to ensure consistency between England and Wales and other constituent countries of the UK. However, this would need to be revised if National Register Scotland (NRS) were to adopt a similar approach in future.

Figure 4.2  
Method for allocating IPS migrant flows to broad stream



**Notes:**

1. All those aged 17 and under and 60 and over are classified as 'Other', apart from 16 year olds and over sixties arriving for 'work' reasons
2. All migrants returning home to live are classified as returning.
3. When distributing IPS those stating 'immigrating' or 'not stated' are allocated to the 'other' reason for visit if there is no data for the relevant age group in the 'work/study/other' categories.
4. Most Asylum Seekers are not captured in the IPS. Home Office data is used to estimate Asylum seeker flows and IPS data.
5. Further splits are made into: i) Non-UK born first-time migrants, ii. Non-UK born returning migrants, and iii) UK-born returning migrants. This is covered further in Section 7.

Table 4.1 shows the final results of the IPS allocation process for mid-2010. These data include adjustments for both visitor and migrant switchers, but do not include asylum seekers. The table shows that some of these splits are small and so the sampling error around these estimates will be large.

Table 4.1  
**IPS splits for the improved methodology, mid-2010**

Category	First-time migrants	Returning		Total
		UK-born	Non-UK born	
Children	20807	5465	2483	28755
Workers	145130	30642	29302	205074
Students	195384	2877	9830	208091
Over 60	3040	5051	1286	9376
Others	28349	13100	4593	46041
All	392709	57134	47494	497337

Work was undertaken to test whether splitting the IPS in this way affects the robustness of the final LA level estimates. The concept was to use the survey estimates and the standard errors to simulate an alternate set of LA level estimates, which could then be compared back to original estimates. This involved using the standard error for each split<sup>12</sup> to produce a range of one standard error around the estimate and then selecting a random point within that range as a simulated estimate for that split. These were then compiled to produce a synthetic estimate for each LA. This process was then repeated a number of times and the results compared. This analysis showed that the use of these splits did not have a significant impact on individual LA level estimates and therefore splitting the IPS in this way was a valid approach.

## 5. Method for distributing IPS worker flows

The methodology for distributing ‘first-time’ workers from the IPS to LAs combines two administrative data sources from the Department of Work and Pensions (DWP):

- The Migrant Worker Scan (MWS)
- The Lifetime Labour Market Database (or L2)

The MWS is the same source that is used for the published figures on National Insurance allocations (NINo) for foreign nationals. Although this is a comprehensive dataset, it is not directly comparable to the ONS 12 month definition of a migrant as it also includes short-term migrant workers. The L2 has only recently become available to ONS. It is a 1% sample of all NINos with information about the amount of national insurance activity in each tax year. With support from DWP, ONS has developed a method for differentiating those who are likely to be short-term migrant workers from those who are clearly long-term migrant workers.

In addition, there are some cases of foreign nationals registering for a NINo to claim state benefits (not for working) and also some cases for which there is a considerable lag between first arrival in the UK and registration. The assumption is made that those who register for a NINo more than six months after arrival are not primarily migrating for work, but for some other reason. These cases are removed from both the MWS and L2 data sets.

<sup>12</sup> Visitor and migrant switchers were not included in this analysis.

The MWS is then linked to HESA to identify those student migrants who have also registered to work. These are removed to minimise double counting and to produce the final MWS count to be distributed. The L2 is used to identify the proportion of:

- long-term migrant workers
- long-term migrants (non-workers or not primarily migrating for work reasons)
- short-term migrants

The long-term migrant work proportions are then applied to the final MWS count. To minimise the standard errors of the L2 sample, some LAs are grouped on combinations of NUTS2 and NUTS3<sup>13</sup> geographies and the proportions are calculated from the grouped data.

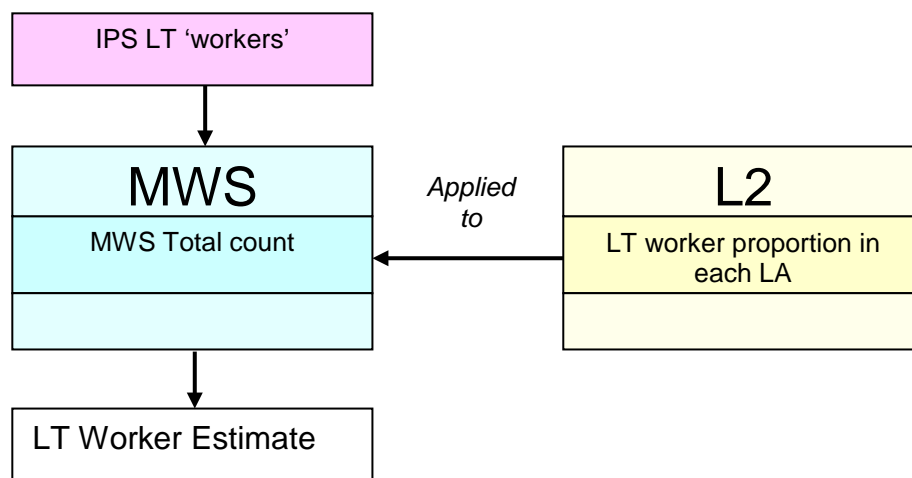
Finally, comparisons between the IPS and L2 show some significant differences for certain groups of countries. Weighting factors are applied to the following set of 'sub-continent' groupings, in order to more fairly distribute the IPS flows:

- EU15/A10 (excluding Cyprus and Malta)
- Asia
- Australia/New Zealand
- Rest of the world.

An overview of the worker distribution methodology is presented in Figure 5.1

Figure 5.1

**Method of Distributing IPS worker flows (first time arrivals) to LA level using L2 and Migrant Worker Scan (MWS)**



## 6. Method for distributing IPS student migrants

The IPS student data is split into Higher Education (HE) and Further Education (FE). These splits are based on IPS data from 2004 and 2005 where these flows can be identified separately. The data for these two years shows that approximately 80% of the long-term study flows are for HE with the rest for FE. This HE/FE split is applied to the IPS total student inflow for all years up until 2010<sup>14</sup>.

<sup>13</sup> The NUTS classification is a hierarchical system for dividing up the economic territory of the EU. For more information see the following [link](#).

<sup>14</sup> From 2011, the HE/FE split will be taken from a new IPS question that was introduced at the beginning of 2011

The student migrants from HESA are defined as those students domiciled abroad in their 1<sup>st</sup> year, who were not known to be previously studying at a UK institution and for whom there is no previous record in HESA or the MWS.

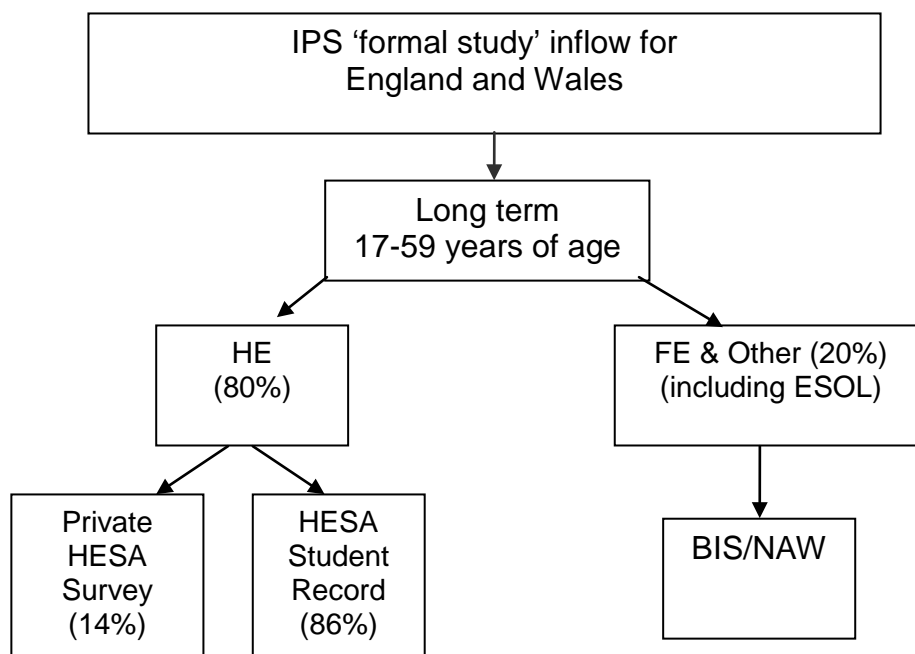
Private universities are not included in the HESA data and are estimated to comprise 14% of the total HE flow. This flow is estimated from a separate HESA survey on private, predominantly HE institutions. Data on FE students come from the Department of Business, Innovation and Skills (BIS) and the Welsh Government (WG). These sources include data on English for Speakers of Other Languages (ESOL) students.

The HESA student record 2005/06 to 2009/10 is used to distribute IPS student HE inflows from 2006 to 2009 between the ages of 17 and 59 to the LA where immigrant students live during term-time. The HESA data only includes the postcode for term-time address since 2007/08<sup>15</sup>. For previous years term-time address is imputed based on the average distribution of term-time LA by campus of the three known years<sup>16</sup> (2007/08 to 2009/10).

BIS/WG data is used to distribute FE/ESOL students aged 16 to 59, allocating them to the LA where they live. Welsh data was complete. For England prior to 2008/09, distributions were imputed based on averaged 2008/09-2009/10 data. Private students' counts from the HESA survey only include campus LA and students are distributed using the term-time address by LA from the HESA record data.

Figure 6.1 provides an overview of the method for distributing long-term student inflows from the IPS.

Figure 6.1  
Method for distributing student flows



<sup>15</sup> In 2007/08 term time address LA was missing for certain campuses. In these cases the 2008/09 distribution was used instead.

<sup>16</sup> Where the university ceased to exist in years after 2005/06 and 2006/07, all students were allocated to the campus LA. This affected less than 200 students/year.

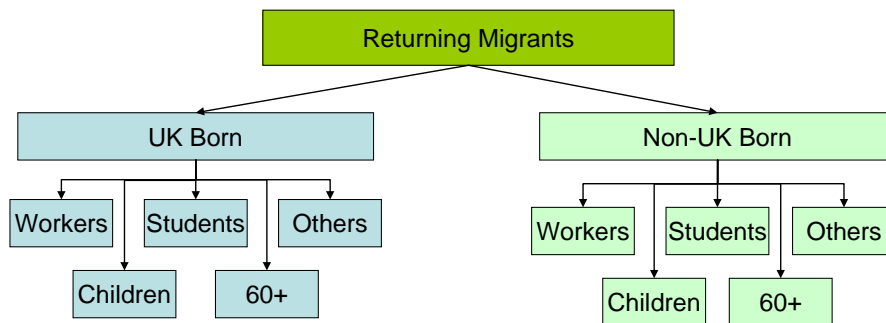
## 7. Method for Distributing Returning Migrants

Returning migrants are identified in the IPS according to whether they have previously lived in the UK. This group is treated separately because many of the administrative sources do not record re-arrivals accurately or consistently. For example, returning migrants would not register with NINo if they had already done so during their previous stay.

Also, analysis has shown that the regional distributions within the UK are different for these two groups. UK born immigrants have a much lower tendency to go to London (approximately 20%) than non-UK born (approximately 40%).

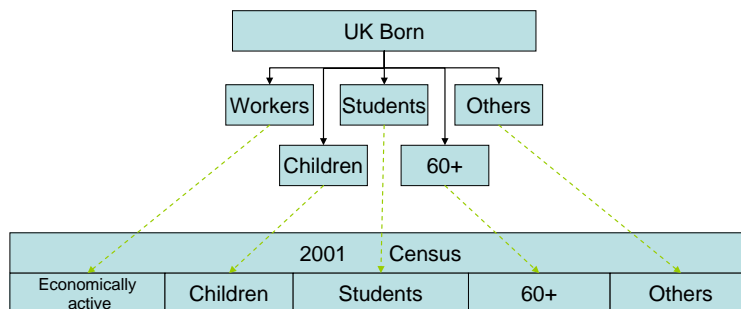
The ‘reason for migration’ groups (shown in 4.2) can be separately identified separately for both UK-born and Non-UK born returning migrants (Figure 7.1).

Figure 7.1  
**Division of IPS returning migrants into ‘reason for migration’ groups**



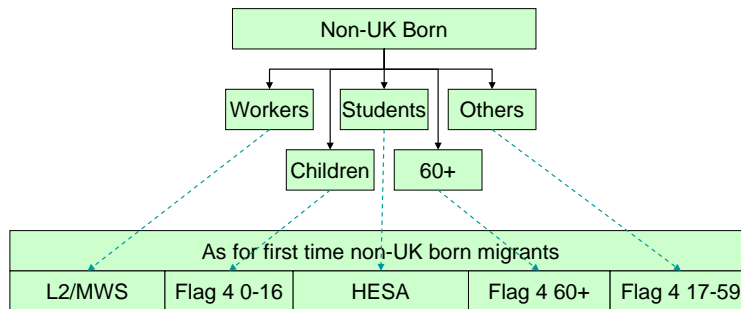
For UK born migrants, 2001 Census data is used to distribute to LA level. The data includes all UK born residents with an address outside the UK one year prior to Census. This data is split by activity (Figure 7.2). The use of 2001 Census data is based on the assumption that the destinations of UK born migrants have not changed significantly over the decade, as regional distributions from IPS and Census are similar.

Figure 7.2  
**Data sources used to distribute UK born returning immigrants**



Regional comparisons of IPS data show that returning non-UK born migrants have similar distributions to first time immigrants, and therefore will be distributed in the same way (Figure 7.3).

Figure 7.3  
**Data sources used to distribute non-UK born returning immigrants**



## 8. Method for distributing ‘other’ migrants

‘Other’ migrants are defined as those migrants whose reason for visit is not ‘work’ or ‘study’. The group is split into three age groups: 0 to 16, 17 to 59 and 60 plus. The 0 to 16 age group refers to all child migrants irrespective of the reason for visit with the exception of 16-year-old ‘workers’ who are included in the worker stream. Similarly for the 60 plus age group, all persons aged 60 and over will be allocated to the ‘other’ group unless they are ‘workers’. These will also be included in the worker flow.

The 17 to 59 group are treated differently. Figure 4.2 shows the reasons for visit that make up this group. There are a number of reasons but the majority are made up of those arriving to accompany or join others.

GP Patient Register data is used to distribute these migrants. When a migrant first registers with a GP they are identified as having a previous address outside of England and Wales, or that they have spent more than 3 months abroad. These people are recorded as having a ‘Flag 4’ in the data and this is considered the most suitable source for distributing flows to LA level.

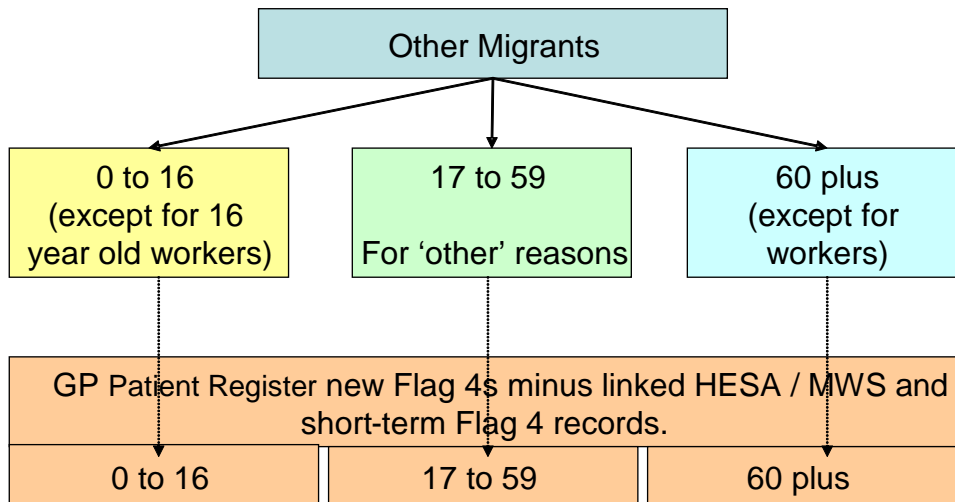
The GP Patient Register data potentially captures all migrants regardless of age and employment status and includes migrants belonging to all four streams (workers, students, other and returning). To minimise double counting the record level MWS and HESA datasets are linked to the record level GP registrations to remove as many workers and students as possible from the GP Patient Register data. This reduces the amount of double counting, although there will always be some workers and students left in the data as not all can be linked.

The longitudinal Patient Register dataset has been created for the period 2005 to 2010. By linking each record by NHS number over six years, we are able to identify a small number of individuals who are likely to be short-term migrants. These too are removed from the Patient Register and this adjusted dataset was used to distribute others.

Further details are covered in the detailed report covering the distribution of the ‘other’ group.



Figure 8.1  
Data source used to distribute 'Others'



## 9. Evidence that the method is an improvement

There are four ways in which the new method can be shown to be an improvement over the current method, namely:

- Increased transparency
- Improved timeliness (in terms of using more timely data that better reflects emerging trends)
- Reduced bias
- Improved accuracy

These are covered in further detail below.

### 9.1 Increased transparency

The elements within the current methodology that determine the ONS immigration estimates for a local authority are complex and include:

- The sparse sample of IPS migrants stating an intention to settle in a particular location.
- The other local authorities making up the intermediate geography (NMGi) with which the local authority is grouped, and the IPS immigration flows into these other local authorities over the last three years
- The regional distribution of migrants identified by the LFS averaged over the last three years.
- The set of co-variables selected by the regression model for each local authority in each year that produce the best fit with the overall pattern of IPS immigration flows to LA level in that year.

These methods and data interact in complex ways which can often make it difficult to relate the data feeding into the methodology to the estimates produced for an individual LA. In addition, the model self-selects the co-variables, some of which do not have an obvious connection with immigration (e.g. 2001 Census data on terraced housing)

Although the improved method does have some elements of complexity, the overall concept of mapping IPS flows to administrative data has a straightforward logic. The improved approach does not rely on statistical relationships between variables, but rather on a logical mapping

between the IPS data by reason for migration and the administrative data in which those migrants should appear. Therefore, the new method addresses the user concerns over the lack of transparency around the current methodology.

## 9.2 Timeliness

The current methodology distributes IPS three year averages at NMGi level and calibrates three year averages of LFS data to distribute at the regional level. The modelling component of the methodology uses a combination of administrative sources and survey data relating to reference period as well as some 2001 Census data. However, the modelling only affects how the resulting estimate is distributed within each NMGi. The bulk of the estimate is driven by the last three years of LFS and IPS data. The resulting lag effects mean that the current method is poor at detecting turning points in immigration.

The sources used in the improved method and how they relate to each reference year are shown in Table 9.1

Table 9.1

### Relationship between administrative sources in the improved method and the reference period

Workstreams	Source	Year				
		2005/06	2006/07	2007/08	2008/09	2009/10
Work	MWS	Estimated from 07/08, 08/09 and 09/10		07/08	08/09	09/10
	L2	05/06	06/07	07/08	08/09	Estimated from 06/07, 07/08, 08/09
	HESA	05/06	06/07	07/08	08/09	09/10
Study	HESA Private	Estimated from 09/10				
	BIS	Averaged from available data	06/07	07/08	08/09	09/10
	WAG	Averaged from available data	06/07	07/08	08/09	09/10
	Others	Flag 4	05/06	06/07	07/08	08/09
Returning Migrants	UK-born	2001 Census				
	Non-UK born	As with Work, Study, Others				

In the earlier years, some sources need to be combined where data is not available. However, for the more recent years, the bulk of the estimate is driven by IPS data for the current reference year and the administrative data for the corresponding period. The exceptions are for:

- UK-born returning migrants, which uses 2001 Census data (although this flow is relatively small and can be shown to be relatively stable over time)
- Private university students, which is based on a 2010 HESA private survey

Together, these two elements comprise only about 15% of the total immigration flow.

The other major issue around timeliness in the improved method concerns the L2. The full L2 data needed to calculate the long-term splits for the current reference period also requires data for the following reference year. For example, mid-2010 requires data for mid-2011, which will not be

available until early 2012. The solution developed is to produce a set of ‘nowcasts’ based on the trend for the previous three years. However, it is important it recognise that this only affects how the MWS data is split and this is available for the current reference period.

Overall, the data used in the improved method is timelier and so will better pick up turning points in immigration. Having timely trends is just as important as the overall level because local authority funding settlement is based on sub-national projections.

### 9.3 Reduction in bias

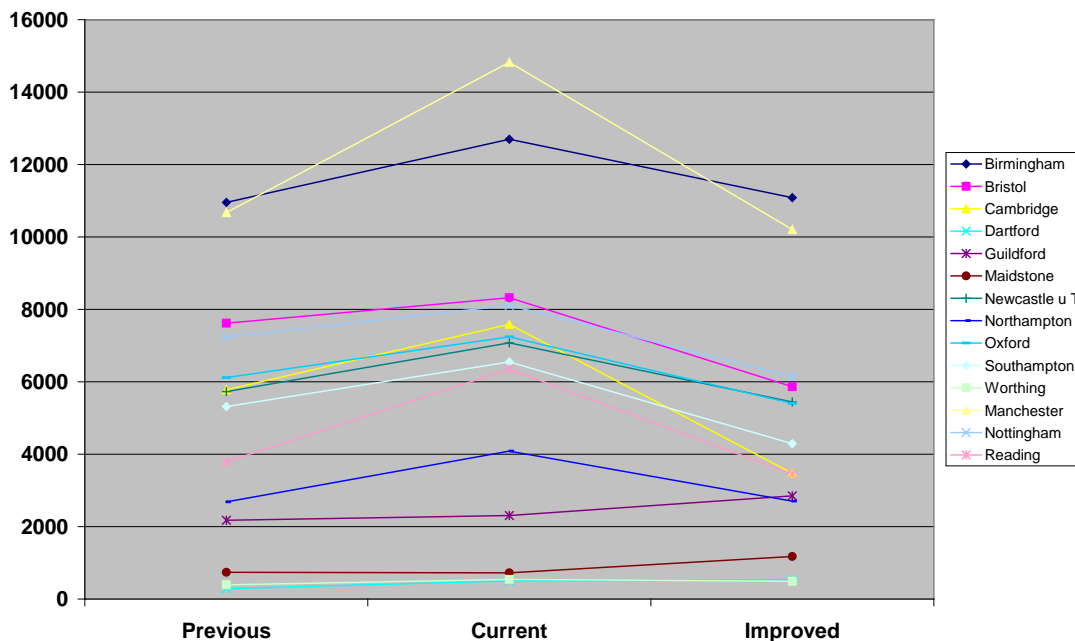
There are two main issues of statistical bias that are addressed by the improved methodology.

#### 9.3.1 Centralising tendency

The concept of IPS centralising tendency was discussed in Section 2.1. There is evidence that the introduction of a model-based approach may have re-introduced some element of IPS centralising tendency.

Figure 9.1 compares immigration estimates using different methods for the fourteen local authorities identified as IPS centralising tendency LAs following the 2001 Census. For the local authorities in this group with higher levels of immigration, there is a pattern of the immigration estimates increasing from the previous approach (i.e. 2001 Census based distributions) to the current methodology (i.e. the model-based approach). However, the pattern moving from the current method to the improved estimates shows a decrease to roughly the levels given by the previous method. This suggests that current method introduced some element of centralising tendency.

Figure 9.1  
Immigration estimates using different methods for selected Local Authorities, mid-2008



During development of the model-based approach, the question of whether modelling the IPS data could result in centralising tendency was explored in some detail (ONS, 2009a). Other

options were explored to address this including modelling at NMGi level, however this did not produce plausible results. Analysis of the modelled estimates at LA level showed that the levels of centralising tendency seen in the IPS data were reduced in the modelled estimates.

Although this earlier analysis does not appear consistent with the results in Figure 9.1, it may be that while the modelling does dampen the effects of centralising tendency of the IPS, it does not eliminate it. The advantage of the improved method improved is that it will not have any IPS centralising tendency bias because it does not use any IPS data about intended destination<sup>17</sup>.

### 9.3.2 Use of LFS data in the current method to distribute to regions

There have been a number of concerns raised about the use of the LFS as the basis for distributing immigration flows to UK countries and English Regions. One issue is the relatively small sample size (typically about 600-700 individual migrant contacts per annum). There are also a number of potential sources of bias including:

- The LFS does not survey communal establishments and so excludes important migrant groups such as international students living in halls of residence; migrants living in hostels/caravans
- Sample clustering (i.e. if all members of a household are migrants they will all be included in the sample)
- Possible response bias in 'hard to count' areas
- Inconsistency with the UN migrant definition (i.e. the LFS identifies migrants based on address 12 months ago being abroad, which could include UK residents living abroad for less than 12 months who would not be counted as long-term migrants by the IPS)

There is evidence that these factors combine to produce regional bias. LA immigration rates per thousand of the population (2009) for the current methodology compared with GP patient register Flag 4 counts are shown for North West (Figure 9.2) and Yorkshire and Humberside (Figure 9.3). In the North West, the Flag 4 counts are higher than the current immigration estimate in 39 out of 43 Local Authorities. In contrast, only 5 out of 21 Local authorities in Yorkshire and Humberside have Flag 4 counts that are higher than the current immigration estimates. This strongly suggests that in 2009, the use of the LFS to distribute immigration to regional level was allocating too many immigrants into Yorkshire and Humber relative to the North West.

The original intention of using the LFS was to tackle the bias resulting from IPS centralising tendency at the national level. Although this may have achieved to a degree, it may have unintentionally introduced other geographical biases into the estimates. This is of particular concern because the LFS itself is weighted up to population estimates that incorporate LFS data. Therefore, there is circularity in the current methodology which will tend to re-enforce any bias effects in the LFS. Since the current method distributes directly from a national level to LA level, there is no requirement to use the LFS.

---

<sup>17</sup> Technically, the LFS remains a small part of the methodology as it is needed to obtain the England and Wales share of the Great Britain IPS immigration estimate, but it is not used below this level.

Figure 9.2  
**Local Authority ranked immigration rates per thousand for Current Method and Flag 4s, North West, 2009**

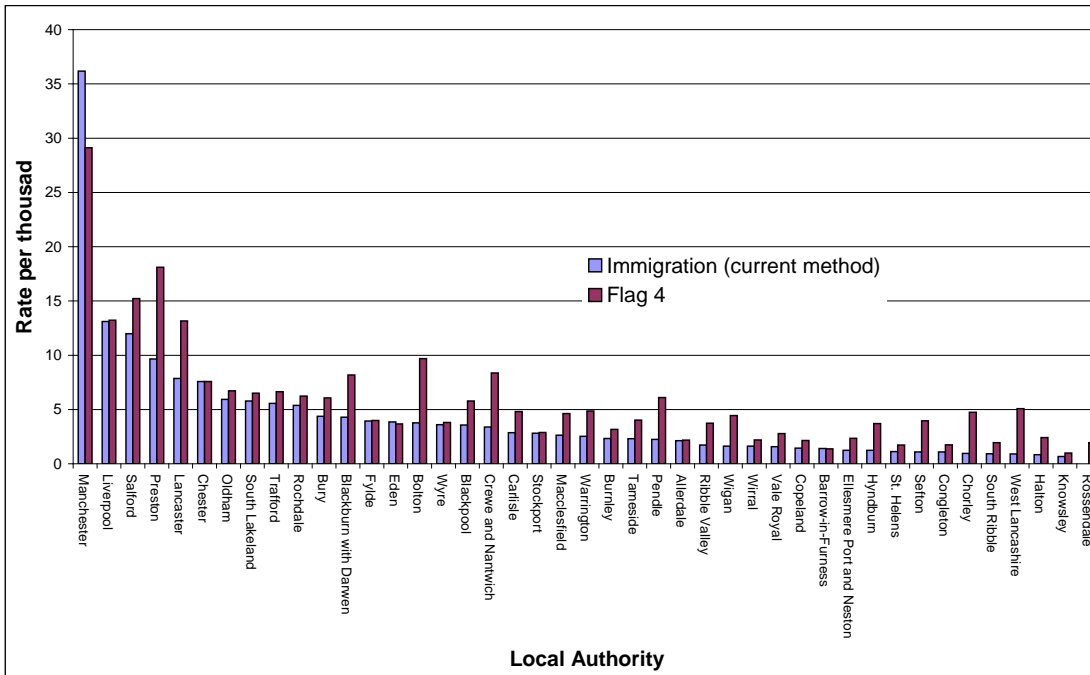
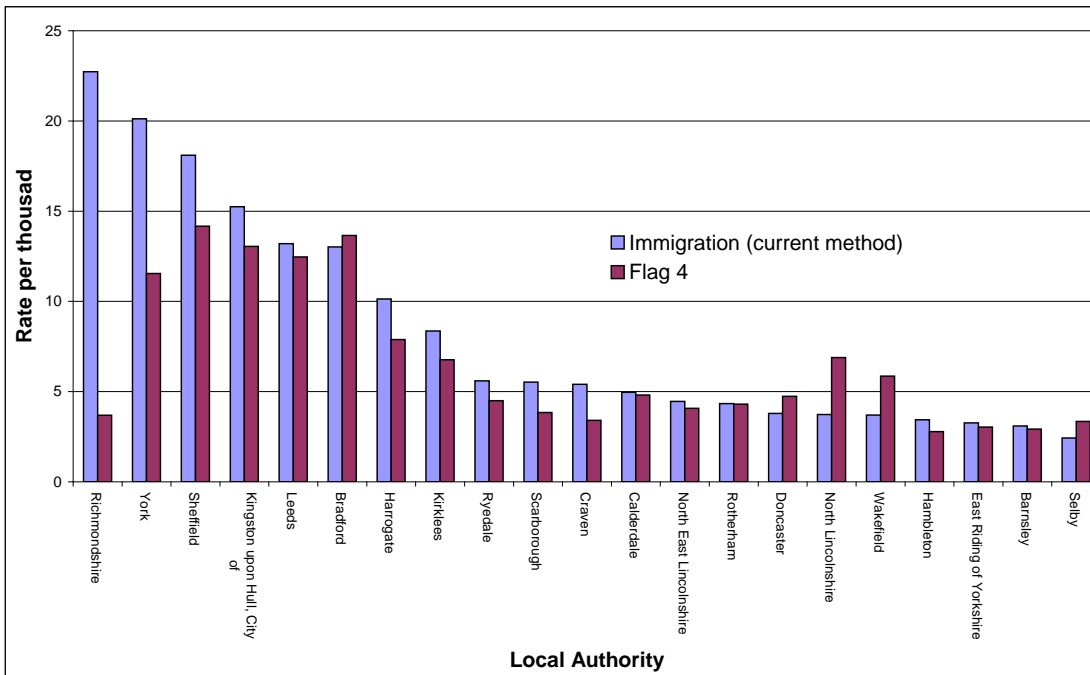


Figure 9.3  
**Local Authority ranked immigration rates per thousand for Current Method and Flag 4s, Yorkshire and Humber, 2009**



### 9.3.3 Minimizing potential bias in the improved method

Section 4.2 discussed possible bias in the improved method due to the issue of double counting. There are also possible bias effects where the administrative data may be over or under-representing certain migrant groups. Steps have been taken to eliminate or minimise these sources of bias. For example, broad country groupings have been used when distributing workers to ensure that the L2 does not over or under-estimate certain flows in the IPS. Data linkage has also been used to minimise the potential bias effects of double counting.

In spite of these efforts, there may be some residual bias effects that can not be corrected for. For example, the use of record linkage to minimise double counting across sources will only work where someone is recorded as living at the same address. This is because all record linking is done on the basis of matching of postcode, date of birth, and gender. If a person changes their address it will not be possible to produce a match. Therefore, there is a possible bias if there are areas where migrants are more likely change address between their initial interactions with the data sources. It is not known whether these patterns vary across areas and so we can not be sure that the proposed distribution method will not introduce some element of geographical bias.

However, any such bias is likely to be much smaller than the bias in the current method. For example, the lack of coverage of communal establishments in the LFS means that up to half of all flows of student migrants are missing from that data. This is just one of several potential bias issues within just one component of the methodology.

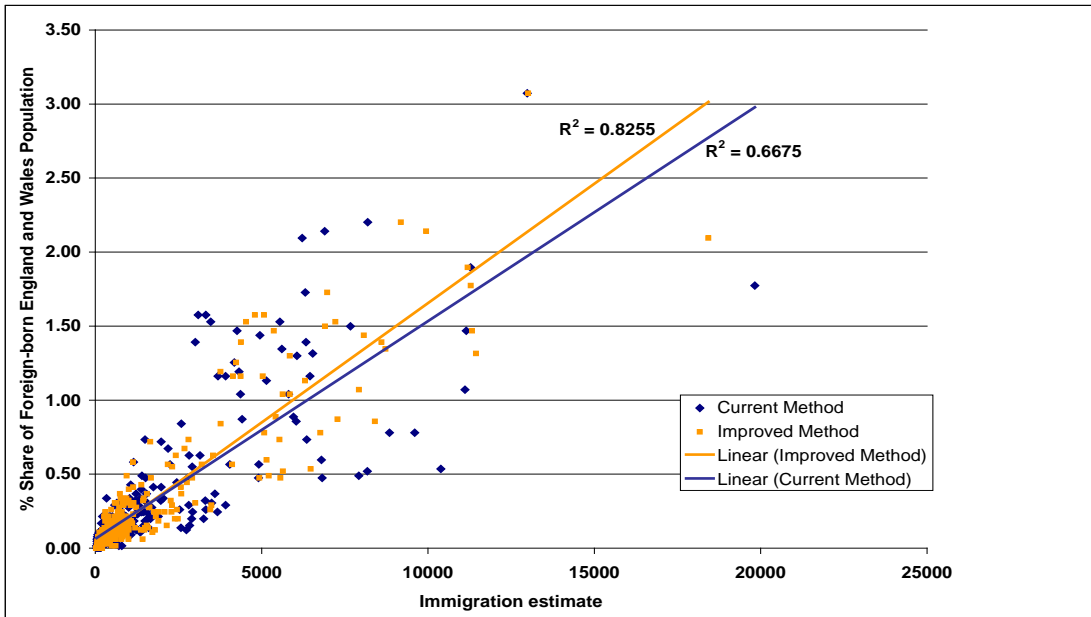
## 9.4 Improved Accuracy

A definitive statement about the accuracy of these estimates can only be made with reference to a gold standard measure, such as a recent Census. The validation of the improved methodology against the 2011 Census results will be a key test of how well the methodology performs. However, even without a gold standard measure, it is possible to evaluate whether the new methodology is an improvement over the current one.

The evaluation approach used has been to correlate the immigration estimates (both the current and the improved) with the local authority share of the England and Wales foreign-born population from the Annual Population Survey (APS). Although there is an obvious link between immigrant flows and stocks, some care is needed with this approach as patterns can vary considerably between areas. For example, student areas will often have high migrant flows relative to their stocks. The APS will also contain some sampling error. Still, this approach has the advantage of being independent from both the current and the improved method and different migration patterns across areas matter less when making relative comparisons.

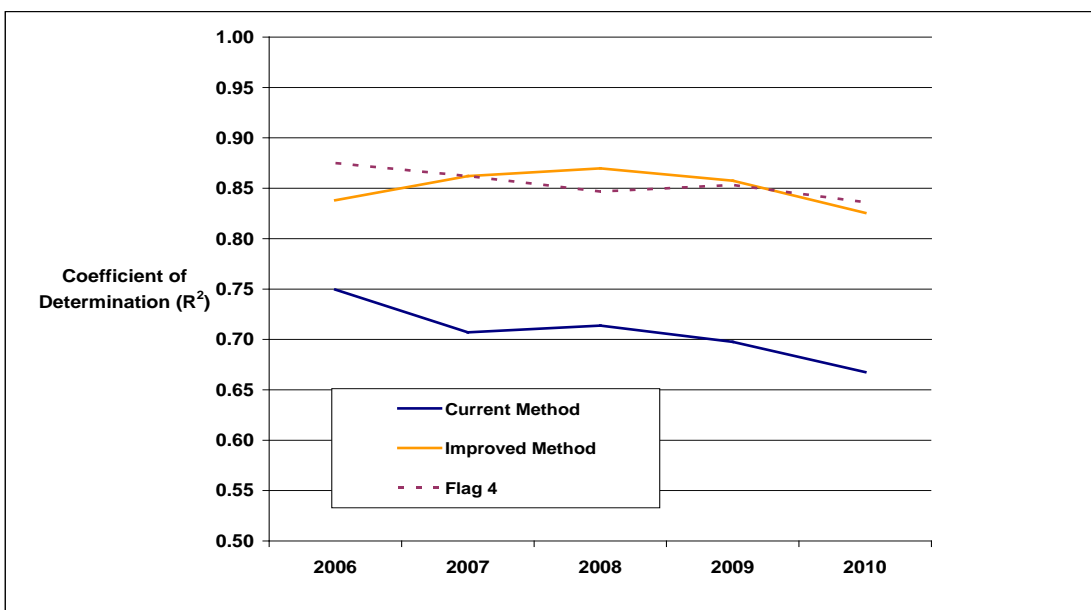
The results of this analysis for both the current and the improved method are shown for mid-2010 in Figure 9.4. The co-efficient of determination (or R-squared value) for the improved methodology is 0.86 compared with 0.67 for the current approach.

Figure 9.4  
**Correlation of immigration (current and improved) with Local Authority share (%) of the England and Wales foreign-born population, Mid-2010**



This set of correlations for all of the five years and the associated r-squared values are shown in Figure 9.5. The equivalent correlations between the GP patient register Flag 4s and the APS foreign-born stocks are also shown. This shows that the improved method gives fairly consistent results over time and compares favourably with the equivalent Flag 4 analysis. The current method does not compare as well and actually seems to deteriorate over time. This very strongly indicates that the improved method is fundamentally better at distributing immigration flows to local authorities than the current methodology.

Figure 9.5  
**Correlation of immigration (current, improved, Flag 4s) with Local Authority share (%) of the England and Wales foreign-born population, Mid-2010**



## 9.5 Conclusion

The improved methodology can be demonstrated to be an improvement over the current method in terms of increased transparency, improved timeliness, reduced bias, and improved accuracy. The underlying strength of the improved method is to maintain consistency with the UN definition of a long-term migrant while exploiting the richness of detail available in the administrative data. This is achieved by continuing to use IPS data as the basis for immigration totals and then mapping this as closely as possible to the relevant data within the administrative sources.



## References:

- Bijak J (2010) *Independent review of methods for distributing International Immigration Estimates to Regions*. <http://www.ons.gov.uk/ons/guide-method/method-quality/imps/updates-and-reports/historical/updates-and-reports-from-2010/independent-review-of-methods-for-distributing-international-immigration-estimates-to-regions---january-2010.pdf>
- Boden P and Rees P (2010) *Using administrative data to improve the estimation of immigration to local areas in England*, *Statistics in Society – Series A, Volume 173 Issue 4m, p707-731, October 2010* Available at: <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2009.00637.x/abstract>
- ONS (2007) *A Review of the Potential Use of Administrative Sources in the Estimation of Population Statistics*, Office for National Statistics. <http://www.ons.gov.uk/ons/guide-method/method-quality/imps/archive-material/archive-updates-and-reports/2007/a-review-of-the-potential-use-of-administrative-sources-in-the-estimation-of-population-statistics.pdf>
- ONS (2009a) *Estimating international long-term immigration by local authority (LA)*, Office for National Statistics. <http://www.ons.gov.uk/ons/guide-method/method-quality/imps/msi-programme/communication/improvements-mid-2008/methodology-papers/immigration-detailed-methology-update.pdf>
- ONS (2009b) *Potential Use of HESA data to improve the regional distribution of international migrant inflows*, Office for National Statistics. <http://www.ons.gov.uk/ons/guide-method/method-quality/imps/msi-programme/communication/improvements-mid-2008/indicative-results/detailed-methodology-papers/use-of-hesa-data-for-the-regional-allocation-of-international-migration.pdf>
- ONS (2009c), *Feasibility report on the potential use of record level Migrant Workers Scan data to improve migration and population statistics*. <http://www.ons.gov.uk/ons/guide-method/method-quality/imps/updates-and-reports/historical/2009/initial-feasibility-report---october-2009.pdf>

## Annex A: Preparing IPS splits to be distributed

Step 1: Split long-term IPS immigrant flows into:

- Work, Study and Other<sup>18</sup> categories (this is detailed in Figure 3)
- Age: Children (0-16), Working age (17-59), and over 60. (Working age is 16-59 if reason for visit is work)
- Citizenship: British, EEA27, Non-EEA<sup>19</sup>

Step 2: Calculate the proportion in each group that are returning migrants.

Step 3: Split visitor switcher pool following the same method as in step 1.

Step 4: Constrain IPS data for long term migrants to calibrated England and Wales total<sup>20</sup>

Step 5: Apply existing migrant switcher fractions to constrained IPS data to remove migrant switchers from the IPS total.

Step 6: Remove returning migrants<sup>21</sup>.

Step 7: Convert IPS data from half year to mid year.

Step 8: Apply existing visitor switcher fractions by citizenship group to visitor switcher pool to produce the visitor switcher adjustment.

Step 9: Convert visitor switcher data from half year to mid year.

Step 10: Constrain summed visitor switcher totals to published totals.

Step 11: Add visitor switcher adjustment to the long-term, first-time migrant IPS total from which migrant switchers have been removed.

Step 12: Constrain totals for new and returning migrants to the Population Estimates Unit (PEU) published totals (excluding the Asylum Seeker adjustment)<sup>22</sup>.

---

<sup>18</sup> The method used to group long-term migrants and the visitor switcher pool into the three groups 'work, study and other' split by new and returning could be simplified. The current method uses both half year and mid year data which may introduce some inconsistency.

<sup>19</sup> The citizenship split is carried out in order to align with the current method for estimating visitor switchers.

<sup>21</sup> Returning migrants can not be identified from IPS visitor data, which is the basis on which on which visitor switchers are calculated. Therefore is assumed that visitor switchers are always first-time migrants

<sup>22</sup> This is needed partly because of slight differences between the Migration Statistics Unit (MSU) and PEU LTIM estimates and to take account of flows from the Republic of Ireland which were treated separately up until 2008.