United Nations ECE/CES/2014/45

Economic and Social Council

Distr.: General
21 February 2014

English only

Economic Commission for Europe

Conference of European Statisticians

Sixty-second plenary session
Paris, 9-11 April 2014
Item 4 of the provisional agenda
Migration statistics

Improving local area estimates of international migration using administrative data

Prepared by Office for National Statistics of the United Kingdom

Summary

This paper explains the challenges in combining survey data, census data and administrative data to create local authority level estimates of international migration. It explains the legal challenges, the data quality challenges, data matching and linking issues and quality assurance issues.

During the first decade of the 21st century the United Kingdom experienced a significant shift in patterns of international immigration which led to a fundamental review of the sources and methods used to estimate net international migration. This led to changes in the International Passenger Survey which resulted in improvements to the national estimates of international migration. However, these changes could not lead to sufficient improvements to local level estimates of international migration. The Office for National Statistics, therefore, developed methods to use administrative data to distribute the national survey estimates to local authority level.

The paper is presented for discussion to the Conference of European Statisticians seminar on migration statistics.

I. Introduction

1. Information about the population underpins some of the key policy and resource allocation decisions that are made by government, business and the voluntary sectors. These decisions affect the economy, the environment and society and have an impact on every individual in the country. It is therefore imperative that the population statistics underpinning these decisions are fit for purpose and make use of the most up to date data that are available from the Census, surveys and administrative data.

2. A key component of the population estimates is migration. During the 1980s and 1990s the level of migration to the United Kingdom (UK) was relatively stable. The methods used to estimate migration over this period and into the first decade of the 21st century relied on using census data and the International Passenger Survey. This method worked well when the patterns of migration did not change from census to census. However, during the first decade of the 21st century, the UK experienced a significant shift in patterns of international immigration which highlighted the limitations of this approach and resulted in several local authorities challenging their population estimates which had been calculated using these migration estimates.

3. As a result, the Office for National Statistics (ONS) established the Migration Statistics Improvement Programme which included the following objectives:

(a) Improve the International Passenger Survey;

(b) Maximise the use made of administrative sources in estimating local authority immigration (both internal and international);

(c) Develop a strategy for improving the information on those entering and leaving the country using administrative sources such as e-Borders (the system that collects carrier passenger data of those travelling to and from the UK) and visas.

4. This paper considers the response to these objectives and in particular how ONS has maximised the use of administrative data sources in estimating local authority immigration.

II. Improvements to the international passenger survey

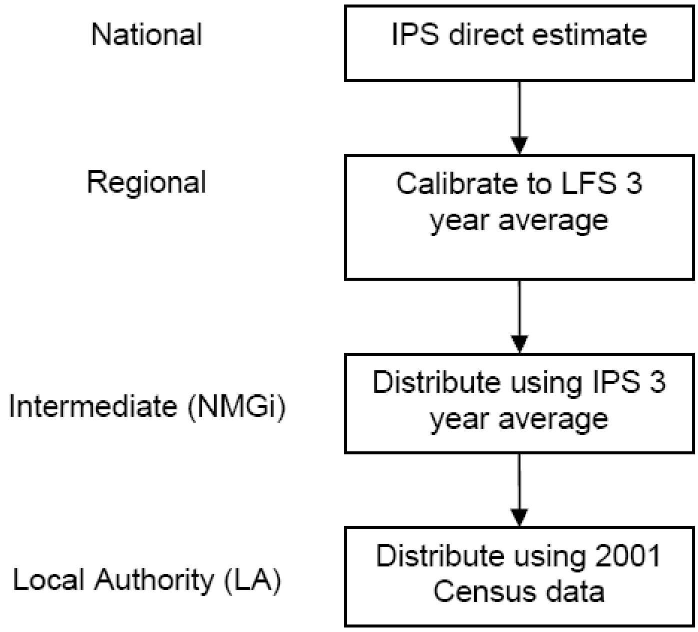
5. Several changes were introduced to the International Passenger Survey (IPS) design in 2009 to better reflect changing patterns of migration to and from the UK, particularly with the accession of eight new countries to the European Union (EU) in 2004. These included changes to the sample size and also the sampling frame which led to increased coverage at regional airports including Stansted, Luton and Manchester. These changes have led to more robust estimates of international migration with lower levels of uncertainty at both the national and regional level. These improvements were vital to subsequent improvements to migration estimates especially to the estimates of migrants arriving from the EU countries.

6. However, these improvements only related to the national and regional estimates of international migration and did not impact on the estimates of international migration into local authorities.

III. Maximising the use of administrative data

7. Prior to 2011, the method for distributing migrants to local authorities is shown in the following Figure.

Figure 1.
Previous method for distributing immigration to LA level outside of London



8. This method took the National IPS method and used data from the Labour Force Survey (LFS), IPS and the Census to derive local authority estimates. The method was very complex and was not easy to explain to users. Although it used the latest data from IPS and LFS to help distribute data to local authorities it still used a three year average from the surveys which meant that the resulting outputs were not able to clearly identify changes in migration patterns. It also relied heavily on the Census which was getting out of date. In addition, IPS collected information on migrant intentions on where they were going to live. These intentions were sometimes inaccurate as some migrants stated major cities as their destination whereas they actually settled in areas surrounding the cities. This led to a bias in the estimates with higher estimates of migration to cities and city centers and lower estimates of migration to the surrounding areas.

IV. Long-term international immigration estimates by local authority

9. After several years' analysis and research a new methodology was released for estimating long-term immigration to local authorities (LAs) in England and Wales. The new approach used administrative data sources to distribute the England and Wales immigration totals from the International Passenger Survey (IPS) directly to LAs. The approach split the IPS into different streams, mainly by 'reason for migration' (e.g. worker, student, other) and then mapped each stream to the most relevant administrative sources which were used to distribute immigrants to each local authority. For example, workers were distributed using National Insurance Number (NINo) data from the Department of

Work and Pensions (DWP); students were mainly distributed using Higher Education Statistics Agency (HESA) data, while children and some other migrants were distributed using migrant information from the patient register (PR) known as ‘flag 4’ data.

10. This methodology was thoroughly quality assured within ONS and externally by academics and local authorities. The method was widely recognised as an improvement over the modelling methods as well as being more transparent and easier to explain. The estimates were used to create indicative population estimates that underpinned the 2010-based subnational population projections.

11. The main features of the new distributional methodology are summarised in Figure 2:

(a) The key principle was to achieve the closest possible mapping between the IPS and the available administrative data;

(b) The Local Authority estimates were based on distributions and not the actual administrative counts. Thus, the total population estimate for England and Wales does not change;

(c) A distinction was drawn between ‘first-time’ migrants and ‘returning’ migrants because of differences in the way in which they interact with the administrative sources;

(d) Record linkage was used both within and between sources to minimise definitional differences and double counting.

12. The sources used were:

(a) Migrant Worker Scan provides a count of foreign nationals applying for a NINo;

(b) Lifetime Labour Market database (L2) is used to estimate the proportion of the NINo count who are long-term migrant workers;

(c) HESA administrative data are used for distributing publicly funded Higher Education student flows;

(d) HESA survey data are used to distribute private Higher Education flows;

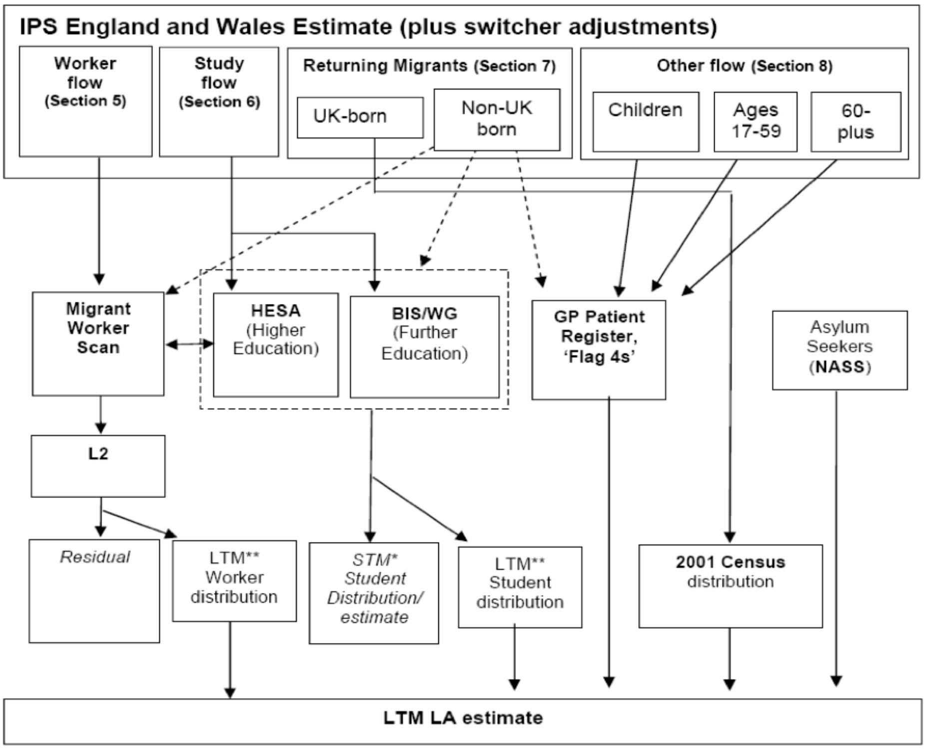
(e) Department of Business, Innovation and Skills (BIS) Welsh Government (WG) are administrative data sources used to distribute Further Education student flows;

(f) 2001 Census data for distributing UK-born returning migrant flows;

(g) National Asylum Support Service (NASS) data to distribute asylum seeker flows identified in the IPS; and

(h) Flag 4 data from the Patient Register Database, to distribute the remaining migrants.

Figure 2.
The main features of the new distributional methodology



*

13. The method was applied for mid-2005/06 to mid-2009/10, partly because this five year period contained sufficient trend data for the 2010-based SNPPs (Subnational Population Projections), but also because some of the administrative data were not available for earlier years. It continues to be used to calculate the current population estimates and projections.

V. Key assumptions

14. The three fundamental principles underpinning the improved methodology were:

(a) The IPS immigration estimate is the best possible estimate at the national level;

(b) 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;

(c) Any differences between the definitions used to map the administrative sources and the corresponding IPS data do not introduce geographic bias.

* 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

VI. Outcome of this approach

15. The approach was, as mentioned earlier, thoroughly quality assured:

(a) Internally by ONS methodologists;

(b) Externally by academic statisticians and demographers; and

(c) Externally by our users.

16. The overwhelming response to the quality assessment was that the method was an improvement reflecting:

(a) That it was transparent and easy to understand. It was intuitively sensible as well;

(b) The administrative data were more timely and would quickly identify changes in patterns of migration.

17. In addition, following the release of the 2011 Census results, a further quality assessment was carried out comparing the results with Census data. It showed that the new method gave closer immigration results in 67% of Local Authority estimates (and many of those that were further away were only slightly so). There were also fewer large differences.

18. There was also a comparison between the Census-based mid-2011 population estimates, and 'rolled-forward' estimates reflecting what the estimates would have been if the 2011 Census data had been unavailable. This comparison found:

(a) 61% of LAs were closer to the Census-based estimates when the new immigration method was used;

(b) The average absolute difference between the rolled-forward and Census-based estimates reduced from 4,500 to 3,800 using the new method;

(c) The key benefit was in the young adult age group (the most common age of migration) - at these ages the new method was, on average, substantially reducing the gap between the rolled-forward and Census-based estimates.

VII. Using administrative data

19. The rest of this paper discusses the challenges and benefits of using administrative data and covers the legal aspects, the quality aspects and the ongoing use of administrative data.

VIII. Legal issues

20. ONS obtained access to a range of administrative sources for statistical purposes, where necessary through the powers that were created in the Statistics and Registration Service Act 2007 or through other existing legal gateways (e.g. Social Security Administration Act 1992). These administrative sources were used either to improve the migration estimates or to quality assure local population statistics and inform future research. In some cases data were linked with other sources to look at how combining sources could inform and improve our knowledge of population and migration.

21. Access to these administrative data sources was a challenge and was successful in part due to the close collaboration between ONS and the departments concerned. Data

from the Department for Work and Pensions (DWP) (Migrant Worker Scan and the Lifetime Labour Market Database), Home Office (Visa and sponsorship data), Department of Health (Patient Register Data), BIS and Welsh Government (Further Education Learner data) and the Higher Education Statistics Agency (Student Records) have been assessed and used in new methods to create local authority estimates of long and short-term international immigration. Data sharing agreements are in place between ONS and these departments to enable the continued supply of these data within the existing legal framework.

22. Special attention has been given to data handling arrangements, as well as to specific legal and privacy issues. ONS has comprehensive procedures in place to protect confidentiality and safeguard the security of all the data that it uses. In addition to complying with agreed Government standards on transfer, storage and handling, individual data sharing applications have been scrutinised as appropriate by Parliament, departmental ethics committees and the Information Commissioner’s Office. All those involved in this work understand the need for, and importance of, safeguarding confidentiality and demonstrating to data owners and the public that ONS can be trusted to handle and use information safely and appropriately.

23. The knowledge and expertise developed in safeguarding confidentiality has been shared within ONS especially with those areas that could have significant reliance on administrative data such as the Beyond 2011 programme which is looking at the future provision of Census type data.

IX. Quality issues

24. ONS carried out a substantial quality assurance assessment for each of the potential administrative sources to establish whether or not the data source could be used. These assessments were wide ranging and assessed the coverage, the timeliness and the completeness of each dataset. This involved looking at areas such as the level of missing data, the level of duplicate records and which variables could potentially be used to link with other datasets.

25. This necessitated close and collaborative working relationships with the owners of the administrative data source and in many instances resulted in substantial improvements in the quality of the administrative source.

26. These assessments were an essential part of the process before an administrative source could be used in the production of the international migration estimates.

X. Ongoing supply and changes to the administrative data sources

27. A key risk in using administrative data is that the coverage and quality can change over time especially when there are changes to the administrative system. Although this risk cannot be eliminated it can be managed through close and regular contact with the owners of the data. This helps to ensure that ONS is given sufficient notice of any changes that are going to be introduced that will have an impact on data quality.

28. ONS has a dedicated team that is responsible for maintaining close relationships with the owners of administrative data sources. This team also ensures that ONS:

(a) meets its obligations under any legislation or data sharing agreements

(b) receives and loads the data into an appropriate and secure IT environment

XI. The future: e-Borders

29. The e-Borders Programme is being implemented by the UK Border Agency (UKBA) with the primary aim of improving UK border security by collecting information from carriers on all those travelling to, and from, the UK. As part of the Border Technology Programme, the UK Border Force collects travel information from carriers, of people travelling to, and from, the UK. The primary purpose of the programme is to improve UK border security; however data from this 'e-Borders system' is also recognised as having the potential to deliver statistical benefits for ONS migration and population statistics.

30. Delivering a strategy of the potential statistical benefits from e-Borders was a key aim of the Migration Statistics Improvement Programme.

31. The e-Borders project highlighted some of the benefits which could be achieved over the medium to long term. It recognised that data from e-Borders will not replace existing sources but will, instead, improve our understanding of migration by combining it with existing sources. Some of the benefits highlighted include improved estimates of international immigration and emigration, improved weighting of the IPS and, through linking, the potential to make further improvements to data on patterns of migration.

32. In 2011, ONS analysed an extract of the e-Borders system data and concluded that the level of coverage was insufficient to deliver the benefits at the time. ONS has recently received a new data extract and are now undertaking a second phase of feasibility research. The aim is to assess whether coverage and data quality have improved sufficiently for ONS to use the data to improve migration and population estimation, and to deliver the potential benefits. Final reporting on this second research phase is planned for mid-2015.

33. A key work strand will involve assessing whether reliable travel histories can be created with the data. Travel histories would be created by linking an individual's travel events over time. Reliable travel histories are crucial if the data are to be used for migration counts or for linkage to other sources.

XII. Conclusion

34. ONS has used a range of administrative sources to improve its estimates of international immigration to local authorities. It has used administrative data alongside survey data to deliver simple and transparent methods that can identify changes to migration patterns in a timely way.

35. The success of this method depended on (1) the existence of legal gateways and the willingness of the owners of administrative data to collaborate and make changes to their administrative data and enter into data sharing agreements and (2) a thorough assessment of the quality of the data source to ensure that it was fit for purpose.

36. There are clearly significant benefits to be gained from using administrative data to estimate migration as we move into the future and the e-borders data has the potential to deliver improvement to the national estimates for both immigration and emigration when used alongside the International Passenger Survey.