A conceptual framework for UK population and migration statistics

Date: 27th March 2012
Coverage: UK
Theme: Population Change

Executive Summary

This report presents a conceptual framework for UK population and migration statistics. It has been produced in collaboration with the Southampton Statistical Sciences Research Institute (S3RI) and the School of Geography at the University of Leeds as part of the Migration Statistics Improvement Programme (MSIP). The aim has been to produce a framework that will facilitate communication with users of population and migration statistics through the development of a shared understanding of the underlying concepts, the available data and the methods used to produce key outputs.

Part I of the report develops the framework.

Population and migration statistics are produced to meet a diverse range of requirements relating to resource allocation, policy making, local service provision, commerce and research. They provide snapshots of population change over time, which is influenced by wider social, cultural and natural environments. This complexity is increased by the range of potential data sources from which population counts may be derived. The development of a framework is therefore essential to provide a structured way to link all the detail involved in translating requirements at a high level to a coherent set of outputs.

The framework contains the key ingredients of concepts, data, processing and estimation and outputs. It is presented in Figure A overleaf. The concepts are the aspects of population and migration of interest; the framework highlights that information about location and time is also required to provide a complete definition. Data are any information gathered via censuses, surveys or administrative registers that may be of use in measuring the concept. Processing and estimation refer to the need to clean, impute, combine data sources and develop statistical models to give a closer representation of the concept than can be achieved with a straightforward count from a single data source. Outputs are the published statistics and are the product of the concepts, data and methods. Understanding of all these elements of the framework is necessary to make sense of how outputs can be designed to meet requirements.

It is vital that the principal population and migration statistics outputs are coherent. One way to achieve this is to measure the population directly and to derive the components of change that contribute to this. The alternative, and the method adopted in the UK, is to measure the components of change and to combine these to derive the population estimate. Within this approach data on registrations of births and deaths are of a high quality. Measures of migration, however, are more problematic. No single source provides the data necessary to accurately measure the concept this model requires. The obstacles include differences in population coverage, small sample sizes and underreporting of migration within the UK.
Part II explains the data sources, processing and outputs currently used in more detail. It covers population estimates and measures of internal and international migration.

It is rare for data sources that may have been designed for different purposes to have identical population coverage. A venn diagram approach provides structure for understanding the differences in coverage between data sources. This can be used to identify how and where processing and estimation are needed to provide coherent outputs that match as closely as possible to the concepts.

This report is just the first step in the development of the conceptual framework for population and migration statistics. The next stage is to use it as the basis for determining priorities as the Migration Statistics Improvement Programme ends. However, there is also a need to develop this work further by exploring how population projections fit, developing the UK angle more thoroughly, exploring migration as a topic more thoroughly and better linking requirements and outputs.

Any queries or feedback in relation to this research report can be done by e-mailing imps@ons.gov.uk

---

1 Categories under “Actual population at time t” are not necessarily mutually exclusive.
Conceptual Framework for

UK Population and Migration Statistics

27 March 2012

James Raymer
Phil Rees
Ann Blake

with contributions from
Peter Boden
James Brown
George Disney
Nikolas Lomax
Paul Norman
John Stillwell
Contact details

Social Statistics, Social Sciences, University of Southampton, SO17 1BJ, UK
James Raymer, j.raymer@soton.ac.uk, Phone +44 (0)23 8058 2935
George Disney, gd1e10@soton.ac.uk
James Brown, j.j.brown@soton.ac.uk

School of Geography, University of Leeds, Leeds LS2 9JT, UK
Philip Rees, p.h.rees@leeds.ac.uk, Phone +44 (0) 113 343 3341
Paul Norman, p.d.norman@leeds.ac.uk
John Stillwell, j.c.h.stillwell@leeds.ac.uk
Nikolas Lomax, eng4nml@leeds.ac.uk

Edge Analytics, Leeds Innovation Centre, 108 Clarendon Road, Leeds, LS2 9DF
Peter Boden, pete@edgeanalytics.co.uk

Migration Statistics Improvement Programme, Population and Demography Division, Office for National Statistics, Segensworth Rd, Titchfield, Hampshire, PO15 5RR, Room 2300
Ann Blake, ann.blake@ons.gsi.gov.uk, Phone: +44 (0)1329 444 640

Contract: This report was prepared by the University of Southampton for the Office for National Statistics under a Contract for the Provision of Research Services in Statistical Methodology, Contract Ref PU-10/0141 and under the direction of Ann Blake. The University of Southampton was supported in the delivery of the report through an Academic Collaboration Agreement on Conceptual Framework for UK Population and Migration Statistics between Research and Innovation Services, University of Southampton and Research Support, University of Leeds.

Acknowledgements: The authors would like to thank everyone who has commented and provided feedback on earlier drafts of this report.

A conceptual framework for UK population and migration statistics

TABLE OF CONTENTS

Contact details Ii
Contract Ii
Key words Ii

Table of contents lii
List of tables V
List of figures Vi

1 INTRODUCTION
1.1 Aims of the framework 1
1.2 Users of population and migration statistics 2
1.3 Outline of report 4

PART I: THE CONCEPTUAL FRAMEWORK 5

2 THE CONTEXT
2.1 Introduction 5
2.2 A system of demographic, social, labour and household statistics 5
2.3 The context for population and migration statistics 7
2.4 Summary 9

3 THE FRAMEWORK
3.1 Introducing the conceptual framework 10
3.2 Population measures 11
3.3 Natural increase measures 12
3.4 Migration measures 12
3.5 Geographic classifications 13
3.6 Population and migration characteristics 14

4 ERROR AND UNCERTAINTY
4.1 Sources of error and measures of uncertainty 15
4.2 The challenge of measuring migration 17

PART II DATA, PROCESSING AND OUTPUTS 19

5 DATA SOURCES
5.1 Introduction 19
5.2 Population censuses 19
5.3 Household surveys 21
5.4 Administrative data 22
5.5 Geographic referencing 23

6 POPULATION ESTIMATES
6.1 Introduction 27
6.2 Sources of data 27
6.3 Estimation methods 33

7 INTERNAL MIGRATION ESTIMATES
7.1 Introduction 38
7.2 Sources of data 38
7.3 Annual mid-year estimation method 41
## LIST OF TABLES

<table>
<thead>
<tr>
<th></th>
<th>Table Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Topics covered in Richard Stone’s <em>Towards a System of Social and Demographic Statistics</em></td>
<td>6</td>
</tr>
<tr>
<td>5.1</td>
<td>Household questions H1 and H2 on usual residents, 2011 Census, England and Wales</td>
<td>20</td>
</tr>
<tr>
<td>A.1</td>
<td>The structure of a movement population account for all ages</td>
<td>69</td>
</tr>
<tr>
<td>A.2</td>
<td>The structure of movement population accounts for a typical period-cohort</td>
<td>73</td>
</tr>
<tr>
<td>A.3</td>
<td>The structure of movement population accounts for the new-born period-cohort</td>
<td>73</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Concepts, data, processing and outputs</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>The continuous process and contexts of producing population estimates</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Structure of the report</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>The context for population and migration statistics</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>A system of population statistics and flow set within a group, infrastructure and locational framework</td>
<td>8</td>
</tr>
<tr>
<td>3.1</td>
<td>Conceptual framework for population and migration statistics</td>
<td>10</td>
</tr>
<tr>
<td>3.2</td>
<td>A framework for estimating internal and international migration statistics</td>
<td>13</td>
</tr>
<tr>
<td>4.1</td>
<td>Total survey error framework</td>
<td>15</td>
</tr>
<tr>
<td>6.1</td>
<td>Main sources of data that can be used to inform UK population estimates</td>
<td>27</td>
</tr>
<tr>
<td>6.2</td>
<td>Relating administrative sources to the resident population in area j at time t</td>
<td>28</td>
</tr>
<tr>
<td>6.3</td>
<td>Relating the patient register to the usual resident population in area j at time t</td>
<td>29</td>
</tr>
<tr>
<td>6.4</td>
<td>Relating the Lifetime Labour Market Database (L2) to the usual resident population in area j at time t</td>
<td>30</td>
</tr>
<tr>
<td>6.5</td>
<td>Relating the English School Census to the usual resident population in area j at time t</td>
<td>31</td>
</tr>
<tr>
<td>6.6</td>
<td>Relating HESA data to the usual resident population in area j at time t</td>
<td>32</td>
</tr>
<tr>
<td>6.7</td>
<td>The ONS method for producing population estimates and main sources of information</td>
<td>33</td>
</tr>
<tr>
<td>6.8</td>
<td>The demographic accounting steps ONS uses to estimate subnational populations</td>
<td>34</td>
</tr>
<tr>
<td>6.9</td>
<td>A time-space diagram illustrating different migration measurement concepts</td>
<td>35</td>
</tr>
<tr>
<td>6.10</td>
<td>A time-space diagram illustrating different migration measurement concepts including births and deaths</td>
<td>36</td>
</tr>
<tr>
<td>7.1</td>
<td>Main sources of data that can be used to inform UK internal migration estimates</td>
<td>38</td>
</tr>
<tr>
<td>7.2</td>
<td>Relating the Patient Registration Data System population transitions to all moves in or out of area j for a one year period</td>
<td>39</td>
</tr>
<tr>
<td>7.3</td>
<td>The ONS method for estimating annual flows of internal migration by age, sex and local authority in England and Wales</td>
<td>41</td>
</tr>
<tr>
<td>8.1</td>
<td>Sources of data that can be used to inform UK international migration estimates</td>
<td>46</td>
</tr>
<tr>
<td>8.2</td>
<td>Relating the International Passenger Survey’s migration data to all migrants arriving or leaving the UK for at least one year</td>
<td>48</td>
</tr>
<tr>
<td>8.3</td>
<td>Overview of method for estimating immigration flows at local authority level</td>
<td>52</td>
</tr>
<tr>
<td>9.1</td>
<td>Summary of key variables collected from censuses and surveys</td>
<td>55</td>
</tr>
<tr>
<td>A.1</td>
<td>Four age-time plans used with population statistics</td>
<td>71</td>
</tr>
<tr>
<td>A.2</td>
<td>Age-time diagrams showing how the period-cohort estimates work</td>
<td>72</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1 Aims of the framework

This project represents a collaborative effort between the Office for National Statistics (ONS), led by Ann Blake, the Southampton Statistical Sciences Research Institute (S3RI), led by James Raymer, and the School of Geography at the University of Leeds, led by Phil Rees. The aim of this study is to develop a conceptual framework to support the understanding and publication of population and migration statistics in the United Kingdom (UK).

According to Laux (2002, p. 485), a conceptual framework is defined as “a set of organising principles which support the compilation and presentation of a set of statistics. The principles relate to:

- the concepts and definitions underpinning the statistics;
- the sources and methodologies used to derive them;
- the structure and tables used for presenting them; and
- links with other areas of statistics.”

A conceptual framework for UK population statistics is needed for facilitating communication with users of population and migration statistics and for ensuring that everyone understands the underlying concepts, the nature of the data available and the methods used to derive estimates of key statistics.

This project contributes to the objectives of the Migration Statistics Improvement Programme (ONS 2012a), a cross-government programme set up to improve the methods and sources underpinning population and migration statistics. The remit of the Migration Statistics Improvement Programme is to ensure estimates are more relevant to user needs, as accurate as possible and recognised as being definitive and authoritative statistics.

The conceptual framework for population and migration statistics is meant to address the needs of both users and data providers within the UK, the European Union (e.g., Eurostat, European Commission) and the international community. In the UK, annual population estimates by age and sex are produced for local authorities for the allocation of national resources and for the planning of local communities. Population estimates are also used at the international level to compare relative fertility levels, proportions foreign-born and general population health. At the European Union level, there are legal requirements for countries to provide demographic and migration statistics for the development of policies and for general comparisons of demographic change and well-being. EU proposals on migration statistics mainstreaming are designed to improve the data available to monitor migrant integration and the socio-economic situation of migrants.

In the first stage of this project, we reviewed the literature on conceptual frameworks and population and migration data. This review identified the main areas of importance to include in a conceptual framework of population and migration statistics, explained how demographic statistics are collected and used to produce migration statistics and population estimates and demonstrated that a deep knowledge of the meaning of the migration data derived from each source was needed in order to choose the right approach for estimating the population. Finally, we looked at some of the key factors that need to be understood in putting together a framework, including the role of time (and duration), the role of geography (and space), the need to provide measures of uncertainty (confidence intervals) and alternative approaches to integration. We pursue all of these themes in this report, in which we deliver a design of a Conceptual Framework for Population and Migration Statistics that will assist users of these data to understand how the different sources of information complement each other, as well as how they are processed, and how population and migration statistics outputs relate to the framework and each other. It also provides a basis for evaluating the benefits of potential new or alternative sources.
1.2 Users of population and migration statistics

The many and varied uses of population and migration statistics are what drive the production and estimation of them. In this section, we briefly describe and provide examples of five broad types of use, which include resource allocation, policy making and monitoring, local service provision and planning, commercial and research.

Resource allocation. Population and migration statistics are used by national governments to distribute money across local governments in the UK for managing services, such as healthcare and education. They can also be used to provide arguments for increasing or decreasing taxes or for providing certain subsidies. Recently, for example, there have been concerns about the current funding scheme for social security in the UK, which relies on the current workforce to support current pensioners. As a result of an increasingly healthy and ageing society, the ratio of workers to pensioners has been steadily declining and will soon reach unsupportable levels under the current financial system. One response to these concerns has been to increase the age at which people are able to draw their state pension.

Policy making and monitoring. Population and migration statistics are used on a daily basis by policy makers, politicians and public health officials to help inform important decisions (Boyle and Dorling 2004). The statistics are used to establish the boundaries for political constituencies and to provide denominators for other measures, such as fertility or unemployment rates. In public policy, population statistics are used to monitor the impact of policy change. Migration statistics are used to identify types of migrants coming to particular areas and their relative well-being, which is important for policies on integration and education, as well as the control of immigration.

Local service provision and planning. Population and migration statistics are used by, for example, local or regional governments to make policies or plans for future schools, policing, hospitals or transport infrastructure. The statistics are used to manage and plan for future water, power and sanitation needs. The statistics can also be used to decide the number and types of homes to be built, as well as for designing centres of shopping and leisure activities.

Commercial. Population and migration statistics are used by insurance companies to set their rates. Companies use the statistics to target their goods and services at specific groups of people (Boyle and Dorling 2004) and for workforce planning.

Research. Population and migration statistics are required for understanding society and societal change (Stone 1971, United Nations 1975). The statistics are used to show how populations are changing for the better or for the worse. This includes comparisons with other countries in Europe or for areas within the UK. In public health research epidemiologists and public health officials use population statistics to identify and help explain inequalities in health.

In addition to the uses outlined above population and migration statistics attract a great deal of interest from the media and the public.

One important message that we aim to get across in this report, as illustrated in Figure 1.1, is the tension between the concepts and the outputs of migration and population statistics. Statistical outputs are often the result of matching available data to a particular concept. In some cases, the data are processed or combined with other information to fit a particular concept. Furthermore, the published outputs may not meet all of the needs of various user groups. As this report highlights, there are many requirements and types of population statistics. However, there is no single source of information in the UK that covers such a wide variety of needs. Population and migration statistics bring together data from many sources, all with their relative strengths and weaknesses.
The production of population and migration statistics is further complicated because populations are both dynamic and heterogeneous. As illustrated in Figure 1.2, populations continuously change according to the addition of births (B), subtraction of deaths (D) and addition or subtraction of migrants (M). These processes are influenced by the social and cultural environments, economic environments and natural and built environments in which the populations live, as well the intersections between them (Statistics New Zealand (2012). In order to understand population statistics, one must first realise that they only represent a ‘snapshot’ of a population at a particular time. As it is very expensive and time consuming to collect detailed and accurate data on the whole population, censuses in the UK have taken place only once every ten years. In between census years, surveys and administrative sources are used to fill the knowledge gaps, albeit with less detail and accuracy.
With such a demand and need for information on populations and their migration behaviours. Trusted, independent and robust information about the size, structure and characteristics of a population is seen to be an essential underpinning of a modern society (Statistics New Zealand 2011). Such information is essential for improving the well-being, prosperity and legitimacy of modern democratic institutions and society alike. As such, it is vital that the statistics are not only reliable and robust, as previously mentioned, but also that users understand how the different statistics are compiled, how they relate to each other and what each variable actually represents. To achieve this, population and migration information needs to be publicly available, transparent and understandable. A conceptual framework for UK population and migration statistics provides the main basis for this.

1.3 Outline of report

The report is organized in two parts (see Figure 1.3). The conceptual framework for population and migration statistics in the UK is presented in the first part. Here, the context in which population and migration statistics are set is presented, followed by a presentation of the overall conceptual framework and a discussion of the issues concerning error and uncertainty. The second part of the report focuses on data, processing and outputs. The aim is to provide basic information about the main sources of information and then some details about how population, internal migration and international migration estimates are produced, and the issues surrounding them. We also include a chapter on the important population and migration attributes, beyond age, sex and location of residence, that are needed for understanding and explaining societal change. Finally, we end the report with a summary of the main features of the conceptual framework and comment on new developments and future plans for population and migration estimation.

Figure 1.3 Structure of the report
PART I: THE CONCEPTUAL FRAMEWORK

2. THE CONTEXT

2.1 Introduction

This report focuses on developing a conceptual framework for UK population and migration statistics. However, before this framework is developed, it is useful to show how such statistics are embedded in a larger system. In Section 2.2, the framework developed by Richard Stone that covers a broad range of government statistics is described. The focus is narrowed in Section 2.3 to people and households but embeds them in the housing and workplace infrastructure necessary for life. An essential feature of this system is the need to know the precise location of the system entities, i.e. individuals, households, housing units and workplaces. The chapter ends with some conclusions about how the context is important for understanding the collection of data, processing and estimation of these data, and the resulting statistical outputs.

2.2 A system of demographic, social, labour and household statistics

The vision of a complete system of demographic and social statistics was proposed by Richard Stone, a Cambridge economist and Nobel Laureate, and one of the authors of the system of national economic accounts which help measure the performance of national economies. Widening his perspective in the 1960s and 1970s, Richard Stone defined a system of social and demographic statistics (Stone 1971), which he then refined for the United Nations (United Nations 1975). As shown in Table 2.1, the system covers the whole range of government statistics, starting in Part Two with demographic stocks and flows, moving on to families and households, social class and stratification, income and wealth, housing and the built environment, the use of time, social security and welfare, learning and educational services, employment, health and public order. Stone’s system also introduces in Part Two the concept of the life course, which is so important in contemporary social science research. While the remit of ONS covers only parts of the Stone system, the Government Statistical Service covers most of the system in its outputs.
### Table 2.1 Topics covered in Richard Stone’s *Towards a System of Social and Demographic Statistics*

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part One</strong></td>
<td>The System as a Whole</td>
</tr>
<tr>
<td>I</td>
<td>Introduction</td>
</tr>
<tr>
<td>II</td>
<td>Scope and Coverage</td>
</tr>
<tr>
<td>III</td>
<td>The Design</td>
</tr>
<tr>
<td>IV</td>
<td>The Collection of Data</td>
</tr>
<tr>
<td>V</td>
<td>Social Indicators</td>
</tr>
<tr>
<td>VI</td>
<td>Non-Market Activities</td>
</tr>
<tr>
<td>VII</td>
<td>Models of Human Stocks and Flows</td>
</tr>
<tr>
<td>VIII</td>
<td>Time Budgets and Models of the Allocation of Time</td>
</tr>
<tr>
<td><strong>Part Two</strong></td>
<td>Individual Sequences and Subsystems</td>
</tr>
<tr>
<td>IX</td>
<td>The Content of Part Two</td>
</tr>
<tr>
<td>X</td>
<td>The Size and Structure of the Population, Births, Deaths and Migration</td>
</tr>
<tr>
<td>XI</td>
<td>Family Formation, Families and Household</td>
</tr>
<tr>
<td>XII</td>
<td>Social Class, Stratification and Mobility</td>
</tr>
<tr>
<td>XIII</td>
<td>The Distribution of Income, Consumption, Accumulation and Net Worth</td>
</tr>
<tr>
<td>XIV</td>
<td>Housing and its Environment</td>
</tr>
<tr>
<td>XV</td>
<td>The Allocation of Time and the Use of Leisure</td>
</tr>
<tr>
<td>XVI</td>
<td>Social Security and Welfare Services</td>
</tr>
<tr>
<td>XVII</td>
<td>Learning Activities and Educational Services</td>
</tr>
<tr>
<td>XVIII</td>
<td>Earning Activities, Employment Services and the Inactive</td>
</tr>
<tr>
<td>XIX</td>
<td>Health and Health Services</td>
</tr>
<tr>
<td>XX</td>
<td>Public Order and Safety, Offenders and their Victims</td>
</tr>
<tr>
<td><strong>Part Three</strong></td>
<td>Examples and Applications</td>
</tr>
</tbody>
</table>

Source: United Nations 1975

A system of demographic, social, housing and labour statistics, within which population and migration statistics can be placed, is shown in Figure 2.1. The system has three layers: the top layer is the comprehensive framework developed by Richard Stone. Within this layer we identify one component, the size and structure of the population, births, deaths and migration (Chapter X in Table 2.1) and then expand this to form the second layer, which shows the two population stocks linked by survivors, inflows and outflows and with relationship at each time point to housing and jobs stocks. The population part of this middle layer is then expanded in the bottom layer which provides details of the flows into the current UK population estimation model that generates successive population stock estimates.
2.3 The context for population and migration statistics

In Figure 2.2, a graphical representation of the context in which populations change is shown. On the left hand side of the diagram we have a set of stocks in a particular location with a variety of entities at a time point 1. These stocks then pass through change boxes and experience losses or gains that transform them into the corresponding stocks at time point 2. The gains are represented by plus signs while losses are represented by minus signs. Note, changes can occur both within locations and between locations. For short periods the largest flow consists of the survivors of the time 1 population through to time 2. Gains to the population stock occur through new births and new migrations during the time interval. Losses to the population occur through deaths and migrations during the time interval. How these gains and losses are organised into population accounts and their associated equations is explained in Chapter 6 of the report. At the top of the figure, we emphasize the stocks-flows-stocks sequence in general and identify the sources of data which are used to fill in the various stocks and flows boxes.
A conceptual framework for UK population and migration statistics

It is often useful in planning and policy making to disaggregate the stocks and flows by attribute (as in Stone’s Chapters XII, XVI, XVII and XVIII presented in Table 2.1). In the diagram, we indicate three dimensions of population attributes which are of general interest: demographic (e.g., sex, age, date of birth, marital status, living arrangement), social (e.g., educational attainment, occupation, social class, religion, national background and ethnicity) and economic (activity class, employment/unemployment, income, wealth). Once we disaggregate the population into classes using an attribute variable then the flows become multidimensional: people can move between classes over the time interval. In some cases, this movement is strictly controlled: we move from a younger age to an older age over a time interval. In other cases, such as marital status, the moves depend on the way categories are defined. For example, we cannot move from married to never-married, but we can move from married to divorced and then to married again.

So far, we have been talking about individuals as single entities but, of course, individuals actually live in small groups called households, which share living space and living arrangements (e.g., a kitchen, meals, house-keeping). A small proportion of the population (around 2%) lives in communal establishments in which they may have their own room but food is provided centrally (e.g., boarding schools, military bases, prisons, long stay hospitals and student halls of residence). Another group entity is the family, which consists of individuals who are related by marriage, partnership and through generations (children, parents, grandchildren). Some households contain individuals who are not related such as those that live in a lodging house or who share a housing unit while studying for a degree. Changes in the nature of households over decades, such as the rise in the number of lone person households, have impacts on housing requirements and this forms part of the context for population change.

Changes in the housing market also affect the level of fertility and migration, e.g., when housing supply is constrained, child bearing may be postponed. The middle row of stock and flow boxes in Figure 2.2 count the number of housing unit entities, i.e. the residential infra-structure. We recognise housing units which are occupied by households and communal establishment buildings which are occupied by individuals in communal groups. Note that there is not a one-to-one relationship between families and housing units. Housing units can be shared by more than one
family. Similarly, communal establishments can be shared by more than one communal group. There are also vacant housing units and communal buildings, which exist because of time gaps between selling and buying and exiting and entering rental property. The events which change the housing stock are additions of newly built housing units and demolitions of older units together with conversions of buildings into and out of residential use. Migration of housing units plays a very small role when households occupying mobile homes or caravans migrate or very occasionally when whole houses are transported by road to new locations.

At the bottom of Figure 2.2, we present the population in relation to the work system. Individuals are classified into the self-employed, the employed (by firms or institutions including the public sector) and not employed who can be divided into those seeking work, the unemployed, and those not, the economically inactive. Some of the economically inactive are retired from work. These stocks are subject to inflows and outflows over the time interval. For further information about the labour market framework, see Laux (2002). The labour market context for population change is important because it can be a key driver for migration within the UK and international migration to and from the UK. Additionally the labour market status of immigrants in the UK is an important element in formulating immigration policy, and therefore influences the population change process.

2.4 Summary

Placing population and migration statistics in a wider context is useful for understanding the population concepts we need to measure, data collection, processing and estimation. The resulting statistical outputs may or may not coincide closely with the original concept or requirement, depending on the data and resources available. It is also useful to know that population and migration statistics are a key ingredient for understanding society and societal change and are directly linked to the production of other statistics, such those for the labour market, health and well being and social welfare.
3. **THE FRAMEWORK**

3.1 **Introducing the conceptual framework**

The conceptual framework for population and migration statistics is presented in Figure 3.1. Included in this framework are the main ingredients of concepts, data, processing and estimation, and outputs. Concepts refer to particular types of population or migration statistics, such as usual residents, employees, students attending higher education institutions, persons receiving welfare or persons present in the UK without British citizenship. Data are any information gathered about the population of interest and its movements, usually obtained from censuses, surveys or administrative registers. Processing and estimation refer to data cleaning, imputation, combining two or more information sources through matching or proportioning, and statistical modelling. Outputs are the published statistics.

---

**Figure 3.1 Conceptual framework for population and migration statistics**

The concepts of population and migration may vary depending on the need of the user. However, all types of population can be related to the actual population at time $t$ in location $i$. Likewise, concepts of migration can be related to the movement of all people in and out of location $i$ between

---

2 Categories under “Actual population at time $t$” are not necessarily mutually exclusive.
two time points. To estimate particular populations, therefore, one must consider the types of entries and exits (including births and deaths, respectively) between time points \( t-n \) and \( t \), where \( n \) refers to the width of the time interval (e.g., days, months or years).

There are various data sources that can be relied on to capture particular populations present in location \( i \) at time \( t \) in statistical form. The geographies are obtained from the location attributes linked to these data sources. From the data sources, attributes of the population and migrants also can be obtained. The main attributes of populations are age and sex. However, to understand changes in populations or differences across space, it is useful to also have more information, such as place of birth, education qualification, employment status, occupational group and marital status.

Finally, processing is required to match the concepts to the data. In some cases, estimation is required to combine data or to include auxiliary information. The result of the data processing and estimation is the outputs. Due to time, budget and available data constraints, the outputs rarely consider all the different types of population present in location \( i \) and at time \( t \). More often, they include just one type of population of particular need, such as the population that are considered to be usual residents.

As an illustration of how the conceptual framework may be applied, consider the concept of “all migrants arriving in the UK for at least a year residing in local authority \( j \), 2010.” Next, assume that there are only two sources of information available: migrants captured in the International Passenger Survey (IPS) and National Insurance Number (NINo) registrations. In 2010, the IPS identified 536 thousand persons coming from abroad who intended stay at least 12 months. They were surveyed at the point of arrival and were asked about their destination in the UK. However, because of the relatively small sample size, this information is only reliable at the national level. For the same year, there were 668 thousand new NINo applications recorded for non-UK nationals over the age of 16 years. The information was collected at the time of registration with reliable address information but no information on length of stay. The main advantage of the IPS is that it measures “all migrants arriving in the UK for at least a year”. The main advantage of the NINo registration is that is measures those “residing in local authority \( j \)”. To combine these two sources together and to the concept above, one would have to estimate the number of visitor and migrant switchers (i.e., those changing their intentions) for the IPS data and estimate the lag between arrival and registration and the duration of stay for the NINo data. Furthermore, one would need to separate the non-UK nationals aged 16+ years from the rest of immigrants in the IPS. The resulting output would be “non-UK nationals aged 16+ years arriving in the UK for at least a year residing in local authority \( j \), 2010.” This illustration shows how a framework is needed to match the available data to a concept and the obstacles and limitations that often occur with such a process.

The following sections introduce the main aspects of this framework in more detail, focusing primarily on the concepts required for UK national statistics. These include measures of population, natural increase and migration. Geography and the inclusion of attributes are also discussed. In Part II of this report, we present the data and estimation methods used to produce the UK statistical outputs for population and migration.

### 3.2 Population measures

As illustrated in Figure 3.1, at any point in time, there are several populations that may be measured. Furthermore, the size and characteristics of the population may vary greatly, depending on both the time of day and day of the year it is measured. In the UK, a mid-year ‘usual resident’ or ‘night time’ population (30 June / 1 July) is estimated. A ‘night time’ measure captures the population where it sleeps, whereas a ‘day time’ measure captures the population where it goes to school, work, market or takes leisure. Business travellers and visitors are usually excluded from official and international mainstream statistics on population and migration. Temporary workers may be included in official estimates of short term migrants, but not in the usually resident population.
When constructing a time series of populations, it is important to be clear about the criteria for inclusion or exclusion of individuals. The Office for National Statistics has long adopted the concept of “usual residence” (see Section 5.2) to define the populations estimated on 30 June/1 July each calendar year. Mid-year population estimates begin with the start of decade census of population count and then roll forward year by year, adding and subtracting relevant estimates of vital events and migration flows that change the population.  

3.3 Natural increase measures

Perhaps the most accurate information we have is on the number of live births and deaths for locations in the UK over time. This is because all births and deaths have to be registered by law. Births are published according to the age, sex, birthplace and residential location of the mother. Deaths are recorded for all persons by age, sex and residential location.

3.4 Migration measures

“Migration is a loosely defined process that represents the relocation of people during a period of time that causes them to relinquish the ties with their previous locality” (Raymer and Smith 2011, p. 703). Migration can involve people moving within a country, as well as across international borders. The factors that separate migration from other forms of mobility (e.g., daily commuting, weekend commuting, holiday visits or seasonal moves) are generally distance travelled and length of time spent in the destination (or away from the origin).

Migration data are obtained from general purpose censuses / surveys or administrative registers. The practical measures of migration obtained from these sources often do not coincide with theoretical or contextual definitions of migration (Bell et al. 2002). The reason for this is that, unlike births and deaths, there are no legal frameworks for measuring migration. In practical terms, migration can be defined as relocations between administrative areas and mobility as relocations within areas. ONS does not restrict the spatial scale of residential migration, as any such classification would be arbitrary. Therefore, a residential migration within the same suburb is still a migration which may have relevance for the estimation of the population of very small areas, such as output areas or postal sectors.

In recent decades in the UK, migration has gained importance in driving both subnational and national population change. Internal migration is driven by people seeking employment, or a better physical or social environment to live, depending on what stage in the life course the migrant is at. The drivers of international migration, however, differ by direction. Immigration is often motivated by the demand for low-skilled, low-wage labour, the need to fill specific (high-skilled) labour gaps, education, people seeking to reconnect their families, nationals returning home after spending a period of time working abroad, and people seeking asylum. Emigration is driven by (high-skilled) nationals seeking better employment opportunities abroad and by foreigners returning home.

A framework for demonstrating the influence of internal and international migration statistics on the population (UK Statistics Authority, 2009, p. 51) is presented in Figure 3.2. Here, we see that both types of moves are linked to residence status. It also includes the distinctions between immigrants and asylum seekers (and between emigrants and enforced removals / voluntary departures), demonstrating the importance of understanding migrants’ characteristics and motives. Data are obtained through the activities the migrants carry out, such as enrolling for education, seeking healthcare or social benefits. Thus, migration is measured differently depending on the source. Administrative registers usually capture data suggesting events, such as National Insurance Number registrations or registration with a local doctor or health clinic. General purpose censuses and surveys, on the other hand, usually capture transitions, such as place of current residence by

---

3 At each census, careful consideration is given to the population base used in the enumeration, and this base has changed over recent censuses (see ONS 2004 for a thorough review).
place of residence one-year ago. The International Passenger Survey (IPS) captures information about migration intentions.

Source: UK Statistics Authority (2009), p.51

Figure 3.2 A framework for estimating internal and international migration statistics

The diagram above highlights the lack of a single comprehensive source of data on internal or international migration, and the lack of a straightforward approach to estimation. The data sources and estimation methods for internal migration are presented in Chapter 7. The corresponding information for immigration and emigration is covered in Chapter 8. Note, emigration flows are particularly difficult to measure as the persons are outside the UK. In the UK, there are very few data sources available for measuring this component of population change.

3.5 Geographic classifications

Geographic classifications are fundamental for understanding society and population change. There are many different ways of representing geography, depending on the data source. However, statistics usually come in the form of aggregate units, such as local authority districts, counties or regions in England and Wales or electoral wards and council areas in Scotland. Sometimes, the actual geography is not of interest but rather the area type, such as urban, rural or coastal. Geographic information is important for planning schools, hospitals, workforce, as well as for comparing different spatial patterns of residence according to ethnicity or density.

Geographic referencing is discussed in more detail in Section 5.4. In general, geo-references apply to point locations. These may include points within properties such as the centres of property parcels or the locations of mailboxes. Areas are defined territories which have an external boundary and which are referenced by an alphanumeric geo-code and National Grid Reference (NGR) for the centre. The boundary of an area is approximated by a set of straight lines which form a polygon or set of polygons.
3.6 Population and migration characteristics

Age, sex and geographical location are considered the baseline characteristics required for population and migration statistics. For understanding change or differences between population groups, it is often useful to have more detailed attribute information, depending on the need or users. For example, for those interested in migrant integration, information on the foreign population, their levels of education and their occupations are useful. For those wishing to set migration policy, understanding the reasons or drivers for migration is important. For services provision, information on population health, number of children and economic activity are required. The key population and migration attributes needed for understanding society and sources of data are presented in Chapter 9.
4. ERROR AND UNCERTAINTY

4.1 Sources of error and measures of uncertainty

In this section, the central issues and methods for including uncertainty are highlighted. The aim is to inform users about the inherent uncertainty in any official statistics, as well as problems associated with particular population groups or geographic areas. This information about uncertainty can be used by users to understand the estimates better. It can also be used to inform data providers where improvements can be made to their data collection or estimation methodologies and to evaluate the scale of improvement achieved.

Within the area of survey methodology, the Total Survey Error framework (see for example Groves et al. 2009) provides a useful basis for considering the errors behind data being collected via a survey, and elements of Total Survey Error are also useful when considering data generated by a census or administrative system. Figure 4.1 (based on Figure 2.5, Groves et al. 2009, pp. 48) sets out a framework that shows the processes used in gathering population data and where errors arise.

![Figure 4.1 Total Survey Error Framework](image)

The acknowledgement and inclusion of error and uncertainty in population and migration estimates is an important aspect in the production of official statistics. In 2009, the UK Statistics Authority report ‘Migration Statistics --- The Way Ahead?’ recommended that “ONS should flag the level of reliability of individual local authority population estimates.” As part of the Migration Statistics Improvement Programme, ONS has responded by developing methods for including uncertainty in the England and Wales local authority population and migration estimates (ONS 2012f). They have also developed a table of key uncertainty indicators, which can be used to identify local authorities with characteristics associated with greater likelihoods of uncertainty in their mid-year estimates, such as the proportion of students present.

4.1.1 Sources of uncertainty and bias

Uncertainty and bias are the two main concerns with regard to the accuracy of population and migration estimates. Uncertainty refers to the overall accuracy of the estimates. Bias refers to whether the estimates are overpredicted or underpredicted on average, and is usually calculated...
as the sum of percentage differences. Both measures are required to understand the quality of the estimates being produced.

Uncertainty can come into the population estimation model in many ways. At the onset, the base population taken from, for example, a recent census may contain error. The components of change may contain error. In the UK, internal migration and international migration are considered the most problematic in terms of accuracy. Whereas, registrations of births and deaths are considered highly accurate since they have a very clear legal framework and long history of data collection.

The sources of bias tend to be undercounting of population groups that are less likely to fill in questionnaires or register in an administrative source. These include highly mobile groups, such as students, migrants, homeless persons and the armed forces population, as well as populations living in areas of high deprivation, unemployment or crime. For example, the age profile of internal migration in England and Wales, as measured using data from the National Health Service Central Register (NHSCR), exhibits a markedly lower level for males than females between ages 15-29 years (Raymer et al. 2011a, pp. 80-81); in 2007, there were 55 per cent more females in the 20-24 year old age group than males. If we believe the two levels should be similar for these ages, then overall, it can be said there is a 15 per cent undercount of males or about a seven per cent undercount for all migrants in the NHSCR data. Overcounting may also be an issue in some administrative sources as a result of lags in deregistration.

When producing estimates, we know that accuracy is higher for larger populations as it takes more to influence their population change. If a recent census is used as the base population for estimation, then accuracy should decrease over time as one moves further away from this benchmark. For example, we would expect the 2002 population estimates to be considerably more accurate than the 2010 population estimates because the most recent census occurred in 2001. The availability and inclusion of historical time series data and covariate information should also improve accuracy. However, this may not always be the case as each source of additional information contains error and introducing them into the estimate could result in more potential sources of error.

The main strategies for dealing with uncertainty and bias are simple. First, collect the best data available, including both time series of data and covariate information. Second, be clear about the estimation model and assumptions. Finally, include measures of uncertainty with the estimates so that the user has a sense for the quality of the estimates.

4.1.2 The importance of including uncertainty

It is important for the user to understand the potential sources of uncertainty and bias, and have some quantitative information to describe this. This information can help users understand the range of plausible totals for a particular area and the reasons why their area or group is considered to be estimated accurately or not accurately. Measures of uncertainty can point to directions where methodology could be improved or where additional data should be gathered. They also provide a more realistic picture of UK population change and migration. Furthermore, as ONS continues to improve their methodology for estimating populations, international migration and internal migration, the availability of uncertainty measures could inform the extent to which the accuracy of estimates has been improved.

4.1.3 Methods for estimating uncertainty

In general, there are two ways of including uncertainty in population and migration estimates. The first is to present ‘high’, ‘principal’ and ‘low’ scenarios based on assumptions about the plausible ranges the estimates can take. The second is to incorporate the actual or estimated probabilities. Information that can be used to estimate uncertainty in this way includes empirical data (including time series of historical data), auxiliary data and expert judgements. This second option is preferred because it provides a quantitative figure as a basis. However, it is considerably more
A conceptual framework for UK population and migration statistics

complicated because it involves identifying and incorporating all the main sources of error from all the information used in the estimation.

The accuracy of censuses is usually obtained by conducting a post-enumeration survey in particular areas of the country. Sometimes, two sources of estimates may be compared to identify error or biases. An example of this is the comparisons of 2001 Census counts with alternative counts derived from administrative registers, carried out as part of the post-census validity checks. ONS found a large number of ‘missing’ young adult males in the census compared to what they expected. Another example is a comparison between the 2001 Census internal migration flows with the observed flows from the 2001 NHSCR, which highlights the undercounting of young adult males in the NHSCR data.

As mentioned above, ONS is currently developing methods for estimating uncertainty in the England and Wales local authority population and migration estimates. The methodology, described in Fulton (2011), focuses on the internal migration component, the statistical error of the IPS international migration component and One Number Census error estimates for the 2001 Census base. The plan is to produce evidence-based error distributions for each of these components, which will then be combined using a simulation approach to produce an overall error distribution for each local authority.

4.2 The challenge of measuring migration

Measuring migration is particularly difficult, and deserves special treatment here. First, long distance migration is a relatively rare event in most people’s lives. Second, many data sources are not designed to capture migration specifically; rather migration measures are a by-product that can be obtained by additional analyses of administrative databases. This has implications for measurement (duration, coverage), accuracy and detail. Thus, when considering appropriate data sources to measure migration, there are inevitably a number of obstacles to overcome.

The first obstacle is to overcome the difference between what the available data measure and what is required by the population estimation model (Rees 1985). This requires understanding the relationship between the two and the development of methodological approaches to reconcile the differences. The second obstacle is the small sample size of most survey data sources (e.g., International Passenger Survey and the Annual Population Survey). Because most people do not migrate within a given time period, small sample size leads to estimates with very high standard errors and, consequently, the inability to use sample data for subnational areas without auxiliary information. Two strategies can be employed to mitigate the difficulty. The first is to expand the sample size. The second strategy is to use estimation techniques or alternative data sources in combination with the survey to improve the quality of the estimates for small groups. See ONS (2011g) for a detailed example of this approach in practice.

The third obstacle involves coverage differences between the populations targeted in survey and administrative sources and the coverage needed in the model for generating population estimates or projections. The solution is to be fully aware of the differences and to determine a strategy for adjusting the gap.

The fourth obstacle is the underreporting of migrations (as events), which also manifests itself as excessive lags in the reporting of address changes. Underreporting may occur as there is no legal requirement to report a migration event in the UK, and individuals may be slow to advise administrations of changes of address where either (a) there is no incentive to do so or (b) there is no need for them to access a particular service. The solution to this issue is much more complicated requiring considerable effort to make detailed comparisons of sources.

The fifth obstacle is the misreporting of destinations or origins of migration, depending on the data set involved. Misreporting occurs for several reasons, and is primarily related to survey or census data. Respondents arriving in the UK may be unsure or vague about their destination.
Respondents to household surveys or the census may inaccurately recall where they were living some time ago, particularly if they have experienced high mobility in between. ONS strives to ensure that questions concerning geographical location are answered accurately.
PART II: DATA, PROCESSING AND OUTPUTS

5. DATA SOURCES

5.1 Introduction

The data underpinning population and migration estimates may be generated from censuses, surveys and population or administrative registers through application of the demographic accounting equation (see Section 6.2). In the following sections these three general approaches to data collection are reviewed and some of the issues with the approaches are discussed in terms of the Total Survey Error framework presented in Section 4.1. Geographic referencing is covered in Section 5.5. Note, most introductory demographic texts\(^4\) cover data collection and provide useful starting points for understanding sources of demographic data (see also Stone 1971, United Nations 1975 and Smith et al. 2001).

5.2 Population censuses

In the UK, censuses have been taken every ten years since 1801 (with the exception of 1941). They tend to be very expensive to undertake and, hence, only taken periodically. However, they usually contain a large amount of detail needed for studying population characteristics and change and provide a benchmark for more frequent population estimates. The United Nations (2008) defines a census of a population in its 2008 *Principles and Recommendations for Population and Housing Censuses* as:

“A population census is the total process of collecting, compiling, evaluating, analysing and publishing or otherwise disseminating demographic, economic and social data pertaining, at a specified time, to all persons in a country or in a well delimited part of a country.” (United Nations 2008, p. 7)

A census is essentially a count of every individual at the location where they sleep during the night at the end of the Census Day within the defined territory. Even in that pure form, there are issues with births and deaths during the night so, very specifically, it should be the population at midnight at the end of Census Day.

The alternative allowed within the UN’s principles is a census defined as a count of the usual residents of a country at midnight at the end of the Census Day. At the top level an individual is a usual resident if they have lived or *intend to live* within the country for at least one year. Defining the unique location of usual residence within a country is more problematic and a definition has to be made operational. For example if a child’s parents are separated but have joint custody, one location has to be chosen as the usual residence for the child. Students have both parental domicile addresses (where they live with their parents) and a term-time address where they live while studying at University. Censuses have varied in terms of requirement on the student but in the 2001 and 2011 Censuses, students were asked to record their term-time addresses as their usual residences (as they were resident there a majority of days in the year). Individuals in temporary accommodation, such as homeless shelters, who have no alternative address, are considered to be usually resident where they are staying that night. When the census has a population present base for counting, as in the UK in 1991, it will ask questions on usual residence to allow outputs to be produced on a usual residence base. The Australian Census also uses persons present as their basis for counting. To enable outputs based on usual residence to be produced, information from entry and exit cards are used to adjust for usual residents who are overseas on Census Night.

From 2001, the UK Census (including England and Wales) have used usual residence as the population base for counting and outputs. In 2001 the intention to stay in the country was not explicitly defined but left to the interpretation of individuals. However, for England and Wales in 2011 there is an explicit question on intended length of stay for recent arrivals allowing the census to capture short-term migrants, not usual residents by the UN definition, and those that intend to be, or are already, long-term migrants (see Table 5.1). In addition, the 2011 Census includes information on visitors in households on Census Night but there is no record of which usual residents were away on the night so it is not possible to construct the person present base directly. The 2011 Census also contains information on second residences and their use. This opens the possibility of the estimation of both weeknight populations and place of work daytime populations (see Section 3.4).

Table 5.1 Household questions H1 and H2 on usual residents, 2011 Census, England and Wales

<table>
<thead>
<tr>
<th>H1 Who usually lives here?</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] Me, this is my permanent or family home</td>
</tr>
<tr>
<td>[2] Family members including partners, children, and babies born on or before 27 March 2011</td>
</tr>
<tr>
<td>[3] Students and/or schoolchildren who live away from home during term time</td>
</tr>
<tr>
<td>[4] Housemates, tenants or lodgers</td>
</tr>
<tr>
<td>[5] People who usually live outside the UK who are staying in the UK for 3 months or more</td>
</tr>
<tr>
<td>[6] People who work away from home within the UK, or are members of the armed forces, if</td>
</tr>
<tr>
<td>this is their permanent or family home</td>
</tr>
<tr>
<td>[7] People who are temporarily outside the UK for less than 12 months</td>
</tr>
<tr>
<td>[8] People staying temporarily who usually live in the UK but do not have another UK</td>
</tr>
<tr>
<td>address, for example, relatives, friends</td>
</tr>
<tr>
<td>[9] Other people who usually live here, including anyone temporarily away from home</td>
</tr>
<tr>
<td>[10] No-one usually lives here, for example, this is a second address or holiday home</td>
</tr>
</tbody>
</table>

H2 Counting everyone you included in H1, how many people usually live here?

Source: ONS 2011 Census, Household Questionnaire, England; Note: The numbers in square brackets are added by the authors for reference purposes.

Traditionally, the census is considered the gold-standard population count for the specific point in time, when it is conducted with no errors under the representation component of Total Survey Error. However, following the 1991 Censuses of the UK there was concern that individuals had been missed and that the Census Validation Survey was inadequate to measure under-coverage (non-response) and adjust the outputs (Heady et al. 1994). For 2001, census coverage surveys
followed the main census collection in England and Wales (Abbott et al. 2005) and Scotland (NRS 2003) and the combined data were used to measure coverage (non-response) and adjust the full census outputs5. While this was broadly successful (Local Government Association 2003), subsequent studies (ONS 2004a, 2004b) showed it had not coped in the most extreme situations and additional adjustments were made to the population estimates resulting from the 2001 Census for a small number of Local Authorities in England and Wales.

The basic framework for coverage assessment following the 2011 Census6 is outlined in Abbott (2007) and the main component is again the Census Coverage Survey (Brown, Abbott and Smith 2011). Prior to publication of the main Census results there will be a quality assurance process utilising administrative sources. Therefore, any additional adjustments, beyond those driven from the Census Coverage Survey, will be made prior to the first publication; leading to a consistent starting point for the next decade’s estimates.

Coverage, part of the representation component of the Total Survey Error framework, is not the only error faced by a census. The Total Survey Error framework also includes errors under the measurement component and as the census is a self-completion questionnaire measurement also matters. In the 2011 Census both quantitative and qualitative research was undertaken to attempt to minimise measurement error. Additionally the results of a quality survey following the census will be used to assess the understanding of questions amongst respondents and the accuracy of their responses to those questions.

5.3 Household surveys

The ONS conducts a variety of social surveys throughout each year. The Scottish Government also undertake surveys in Scotland. Typically, a survey has a specific purpose7 and estimates quantities from a clearly defined population. A survey can have coverage errors if the population sampled from is not the population the survey wishes to estimate. For example, the population of interest for the International Passenger Survey (IPS) (ONS 2012b) is all international passengers entering and exiting the UK in a specific year. However, the population that it samples is more restricted because not all legal entry and exit points are included in the sample, although more, such as Southampton Airport, have been added recently. The weighting process does take into account traffic through all ports, and tries to account for times that are not covered.

Any survey will, of course, have a sampling error due to collecting data on only a subset of the population. Typically this sampling error is relatively small when one considers the main estimates of a survey, such as estimates of immigration from the IPS for all passengers in a year, although migrants are relatively rare within the population of all passengers and therefore already difficult to estimate from a survey. However, the sampling error becomes considerable when estimates are required for detailed sub-groups, such as estimates of immigration from the IPS for male passengers by age and destination local authority in a year. In addition, the sampling process does not just result in sampling error. Non-response introduces additional error that originates from both sampling (reduced sample size) and inadequate coverage (non-responders differ from responders). A key tool for tackling both errors in social surveys is to calibrate the survey so it recovers ‘known’ quantities such as the age-sex-regional structure of the population. Calibration might, for example, involve re-weighting the age-sex distribution of the sample to match that in a population source. If these quantities are related to things the survey is designed to measure, say labour market participation, the calibration will reduce both the sampling and coverage type errors. This type of approach is not possible with the migration estimates of the IPS as there is no

5 The full adjustment in 2001 is referred to as the One-Number Census. More information on this can be found at ONS (2003).

6 More details on the methodology for the 2011 Census can be found at ONS (2011d).

7 For example the Labour Force Survey (LFS) measures aspects of labour market participation for the household population.
population level information on the migrant population\(^8\) to serve as calibration constraints. Instead, the IPS calibrates to overall passenger numbers and relies on its own classification of migrants within the survey and a relationship between overall flows and migrant flows to produce the migration estimates. Calibration to population estimates also creates a feedback loop when other surveys such as the Labour Force Survey (LFS) (ONS 2012c) are used to construct population estimates.

Defining the actual population to estimate can also be problematic as populations tend to be dynamic. The census approaches this by defining the population at a specific time-point but a survey like the LFS that conducts its sampling throughout the year cannot. In fact, an annual estimate from the LFS is effectively based on ‘averaging’ across the quarters, which themselves are ‘averaging’ over the individual months, which themselves refer to a specific reference period that differs within the month. Therefore, when we say the LFS estimates the number of unemployed in a period, for example, this is based generally on averaging over time and not the responses for a specific time point or period. For the IPS, this is less problematic as the target population is all passengers in a year so it is necessary to sample across time to represent the population.

Within the Total Survey Error framework, the issues discussed so far all relate to representation component. However, as with the census, once we use questionnaires, either self-completion or interviewer completed, the measurement component needs to be considered. ONS undertakes quantitative and qualitative research in order to understand questions such as – Do respondents understand the concept of a migrant as operationalised within the IPS? Can they reliably answer the questions with respect to intended duration of absence from the UK or stay within the UK?

### 5.4 Administrative data

#### 5.4.1 Introduction

A loose definition of administrative data would be information that is collected about individuals in order to carry out a government task (e.g. payment of benefit to recipients, payment of contributions by participants). As far as the system collecting the data is concerned, the needs of that system drive the structure, nature, and quality of the data collected and, although some administrative data are used to produce official statistics, this is not the primary use. For example, the data on the issuing of National Insurance Numbers (NINos) to new migrants is there to manage the National Insurance system. It is not intended to track the movements of migrants into and out of the country. However, it is clear that in many cases the administrative data generated within such a system can have value outside of that system, provided we carefully consider the quality implications of the method of collection.

Within the representation component of the Total Survey Error framework coverage is an issue, as administrative systems typically do not cover the population we wish to measure: it covers the population relevant to the administrative system. For example, the English School Census data cover all the pupils in the state school system. In other words, it has complete coverage of the system it relates to, but no coverage of children outside that system (e.g. pupils in private schools). Such coverage differences should be identifiable based on the definitions used within a specific system. However, unlike the census, there are rarely coverage surveys to measure these issues.

What can matter more with administrative data is the measurement component of the Total Survey Error framework. In particular, how do the data being collected relate to the underlying concepts we wish to measure (validity) and what measurement process is being used to collect that data (measurement error)? For example, does the ethnicity data collected in the English School Census match to the underlying construct of ethnicity defined for the census? Given we have a valid

---

\(^8\) The e-borders project has the potential to provide that information in the future, which would likely reduce sampling and coverage errors in the IPS for estimating detailed information on migrant flows (Home Office, UK Border Agency 2011).
measurement tool, is the ethnicity data always completed by the parents or do some schools collate the data based on their own observations introducing variable measurement across schools? These two issues would be crucial if considering distributions of children’s ethnicity based on data from the English School Census rather than from the population census or major surveys. In addition, the number of variables collected can be more limited than with censuses or surveys as the inclusion of a variable is driven by the primary administrative use of the data. For example, published health demographic data only register sex and date of birth (and hence age); other basic demographic variables, such as ethnicity, are not considered necessary for the purpose of managing funding within the GP system.

5.4.2 Legislative framework

In order to access administrative microdata for use in population statistics, a legal gateway must exist for it to be shared between the data owner and ONS. Data sharing for statistical purposes is covered by the Statistics and Registration Services Act 2007 (UK Parliament 2007) and the Data Protection Act 1998 (UK Parliament 1998). Best practice in the field of data sharing has recently been clarified with the publication by the Information Commissioner of the ‘Data Sharing - Code of Practice’ (Information Commissioner’s Office 2011).

If a legal gateway already exists allowing the data to be shared there is no need for additional legislation. However if there is no legal gateway, then a data sharing regulation, such as between ONS and the Department for Education relating to the English Schools Census is required. The data sharing regulation should include information regarding the purposes for which the data may be used after it has been shared. Specific details regarding how the data are transferred, its storage, the people who are allowed access to it, and the purposes for which it may be used should also be covered in the service level agreement or memorandum of understanding between the data owner and recipient organisation.

Once the data sharing agreement is finalised, the data are stored, processed and analysed in a secure research environment which complies with the standards defined for the specific data source. Only individuals with appropriate clearance who have been trained in the safe and secure use of identifiable information may access the data. Adherence to the UK Statistics Authority Code of Practice helps to ensure that no information about an identifiable person, household or family is disclosed. Outputs may not be removed from the secure research environment without clearance, which includes a review of the disclosure risk associated with the data or analyses being released. Analyses based on the data may not be published or shared further without reference to the data owner first. Data are retained only in accordance with the standards required.

5.5 Geographic referencing

ONS and the National Records for Scotland (NRS) invest substantial resources to ensure that all National Statistics databases have the best possible geo-referencing of individual and household records. The ONS Geographical Service (ONS 2011c) states that:

“Geography is key to virtually all National Statistics. It provides the structure for collecting, processing, storing and aggregating the data. The framework provided by geography is often the only factor different datasets have in common.”

Geo-references used by ONS and NRS contain one or more of the following elements: the address, the address National Grid Reference, the postcode and the postcode National Grid Reference. Some of the key features of this locational infrastructure are now described.9

9 For further information, refer to Raper et al. (1992: 9-28) on the nature of geo-referencing and identification of geographical entities and Martin (1996: 1-4) on definitions of geographical information systems, which are built on geo-referencing, in relation to socioeconomic data.
5.5.1 Geo-referencing of microdata

The collection and dissemination of population related data requires that all database records are geo-referenced. Individual level microdata which record people or buildings are usually geo-referenced as a point in space using a National Grid Reference as an Easting (x) and Northing (y). Population level aggregate information for geographical areas is usually linked to a discrete zone which may be referred to by name or by an alphanumeric code. A database which contains geo-referenced records can be regarded as the substantive information of a geographical information system, which enables the information to be viewed on a map when linked to digital boundary information.

When individuals respond to a census or survey, or when administrative data on individuals are compiled, an address is recorded. This may include a reference to a particular building (e.g. the person's dwelling) and street name but most often the record refers to a postcode. The UK has had a full coverage of postcodes since 1974 (Raper et al. 1992) and the postcode has become a vital piece of geographically-related information for the collection and processing of population related data.

Microdata records with the postcode of the person's residential address can be linked to geographical areas by matching the individual records to the postcode directory. This represents a major resource to users who have a micro-database which they wish to link to official data sets. Examples of use of geo-codes are: (1) the geo-coding of individual records in censuses and surveys for subsequent aggregation into geographical zones for release, (2) the geo-coding of administrative records (such as births, deaths or benefits registrations) for aggregation into areas for release, (3) linking surveyed databases with census databases to add value for research purposes and (4) linking customer databases to areas to identify catchment areas.

5.5.2 Postcode Directories

Postcode Address File directories are compiled by Royal Mail which include information about the location (e.g. number of delivery points) and a point grid reference. Three aspects should be noted. First, the postcode system has been devised for the sorting and delivery of mail and its alphanumeric codes have specific meanings in this context. Second, whilst a postcode is allocated a point grid reference, in reality it is a list of house numbers (so can be perceived as a line along the relevant street or a polygon drawn around all the properties). Finally, postcodes can change over time in their existence and in their location. The point location is usually referred to as the postcode centroid. The centroid is given the Address-Point coordinates of the nearest delivery point to the calculated mean position of the delivery points in the unit (OS 2010, p.9). Individual records from a census, survey or administrative dataset can be assigned the grid reference associated with the postcode.10

Postcode directories have improved over time through the build up of the historic record of postcode existence (including dates of introduction and termination), the better resolution of the National Grid Reference which thereby increases the spatial precision and the increasing number and variety of the area geographies to which each postcode is linked. Linkages between postcode and area are achieved using geographic information system software and the point in polygon function.

---

10 Following the 1991 Census, a step change away from clerically assigned linkages between microdata and geographical areas occurred with the release of the Postcode-Enumeration District Directory for England and Wales and an equivalent in Scotland. Subsequent postcode directories (which are enhancements of Royal Mail products) include the All Field Postcode Directory (Simpson, 2002), the 2001 Postcode Headcount file (Rees et al., 2005), the National Statistics Postcode Directory (NSPD) (ESRC Census Programme 2011) and the ONS Postcode Directory (ONSPD) (ONS 2011c).
5.5.3 Address directories

The requirements of the 2011 Census for a National Address List for posting out census questionnaires to householders led to the collaboration of the main providers: the Royal Mail (creators of the Postcode Address File), Ordnance Survey Great Britain (creators of the product Address-Point™) and the Local Government Association (developers of the National Land and Property Gazetteer or NLPG). The latter two organizations have now created a public sector limited liability partnership called GeoPlace to create and update the National Address Gazetteer Database and National Street Gazetteer for England and Wales (GeoPlace 2011), supplying definitive publicly-owned spatial address and street data. The partners are negotiating the extension of coverage to Scotland. ONS will, in future, source its geo-referenced address data from GeoPlace (Calder, 2011).

5.5.4 Geo-referencing of area data

Administrative processes drive the collection of population-related data and affect decisions about the geographies for which data are disseminated. Administrative geographies are used to publish census data at local government scale but for the dissemination at sub-local government level, statistical geographies are defined. Output areas based on postcodes, for example, are defined according to the minimum population or household count that is required to help preserve the confidentiality of people’s personal information. Most geographies are hierarchical with smaller units nesting into those which are larger. Alphanumeric coding systems used to identify areas often reflect the hierarchies (see Martin, 1996, pp. 74 and 77).

5.5.5 Geo-referencing institutional populations

Institutions such as care homes, student halls of residence, prisons, armed forces bases or boarding schools present problems for geo-coding, with inconsistent methods used over time and space. Currently, the decennial census is the only survey which covers both residential and institutional populations. In terms of data collection, the residents of an institution represent microdata, but the variation in scale of the institution itself and style of organisation make it hard to define in terms of geographical entity whether the ‘dwelling’ part is a point within the polygon of the land the institution occupies.

Due to these definitional issues, ONS and NRS devote a lot of effort to understanding the location attributes of institutions that house residents. In the postcode directories, there is a field which indicates whether the address relates to a ‘small user’ (assumed to be a residential address) or a ‘large user’ (assumed to be a business). An institution may be recorded by Royal Mail as either of these or even as a non-geographic PO Box. To alleviate these problems, after 2001 all census output is based on a frozen set of small user postcodes, and all small user postcodes are geographically referenced. So if an institution had a large user postcode or a PO Box number then the output is to its small user link and the associated National Grid Reference.

5.5.6 Geo-referencing of migration data

On census questionnaires, respondents are asked to provide their address one year ago if different to the address on the questionnaire. The postcodes of previous and current address allow one year transition data to be computed as interaction data and disseminated as origin-destination matrices in the Special Migration Statistics. The distance of move can also be computed as a straight line distance between the eastings and northings of the postcode points.

Outside census years, estimates of migration between areas are have been produced from health service data on patients who have changed address and who have informed their existing or new General Practitioner. The geographical units for migration data were based on health service geographies up to the late 1990s and local authority districts thereafter. Underpinning the geo-
referencing process is the postcode of patients, whether or not they have changed from one time-point to the next and aggregations of postcodes at origins and destinations to produce interaction matrices. There are broader spatial considerations which must be taken into account when migration statistics are generated and analysed. For internal migration, the volume of migration increases as the number of spatial units increases.

5.5.7 Small area geographies

The taking of censuses in general has been the opportunity to count and define the extent of small area geographies for the safe release of population data. For instance, the design of the small area geography of the 2001 Census resulted in more detailed address level geo-coding used and postcodes being synthetically digitised as polygons (Martin, 2002). Output Areas (OAs) were defined as a statistical geography whereby the zones were small enough to allow detailed geographical research but large enough (in population and household numbers) for the safe release of cross-tabulated socio-demographic data; grouped where necessary. Soon after the 2001 Census further statistical geographies were defined as aggregates of OAs (Super Output Areas in England, Wales and Northern Ireland; Data Zones in Scotland) and these have become the main geographical zones for which post-2001 administrative data are released. For the 2011 Census, the 2001 definitions of Output Areas are being retained as far as possible though population shifts as a result of new developments or redevelopments will necessitate some re-design to ensure threshold sizes for OA populations are met (Cockings et al. 2009, 2011).
6. POPULATION ESTIMATES

6.1 Introduction

ONS' remit with regard to population estimates is twofold – firstly, to produce national and local authority estimates for England and Wales and, secondly, to bring these together with estimates from National Records for Scotland (NRS) for Scotland and from Northern Ireland Statistics and Research Agency (NISRA) for Northern Ireland to create UK level estimates. In this section, we discuss how ONS constructs population estimates, describing the data sources available and presenting an overview of the current rolled forward estimation methodology. More details of these methods are given in the Appendix.

6.2 Sources of data

Population estimates can be achieved in several ways. Estimates may be the result of direct counting from a population register or census. Estimates may be made using counts and adjustments from a set of administrative registers. Estimates may be created by adding and subtracting relevant demographic flows from a prior population count or estimate. As illustrated in Figure 6.1, there are many data sources that can be used to inform UK population estimates. The main sources used are the census and derivations from demographic accounting.\(^{11}\) Other sources of information that could be used to estimate populations include the National Health Service patient registers, the National Insurance Recording and PAYE System (NPS), the English School Census, the Higher Education Statistics Agency data and the Annual Population Survey. ONS is currently exploring the quality and application of these data sources, including combinations of them, to estimating populations as part of their ‘Beyond 2011’ programme and to quality assuring population outputs as part of the Migration Statistics Improvement Programme.

![Diagram showing sources of data](image)

Figure 6.1 Main sources of data that can be used to inform UK population estimates

\(^{11}\) Refer to Section 6.2.
A general description of data sources covering censuses, surveys and administrative records is covered in Chapter 5. The legal gateways for access to the data referenced in this section can be found in Section 5.4.1. Below, we focus on some specific administrative data and survey sources that can be used to directly or indirectly inform annual population estimates by age, sex and geography. Here, it is useful to refer to both the conceptual framework presented in Figure 3.1 and the Total Survey Error framework provided in Figure 4.1. These frameworks can be used to assess how the data sources relate to the concepts we are concerned with, including how they capture information about location, births, deaths and migration, and to determine where potential sources of error may arise, respectively. The assessment is conveyed visually using Venn diagrams as outlined in Fig 6.2 below. The aim of these diagrams is to illustrate how the concept of the usual resident population relates to the population covered by the administrative source in a specific area \( j \) at a specific point in time \( t \).

**Figure 6.2 Relating administrative sources to the usual resident population in area \( j \) at time \( t \)**

### 6.2.1 Patient register

General Practitioners (GPs) are the first point of contact for nearly all National Health Service (NHS) patients (NHS 2011a). Most individuals and households are registered with a GP near to their home. The NHS patient registers are used to maintain an accurate list of all persons registered with a GP, allowing the timely transfer of medical records and correct payments to doctors. They contain a list of everyone who is, or ever has been, registered with a GP in England and Wales since the foundation of the NHS in July 1948 (NHS 2011b). Although this source of individual-level data has a specific administrative purpose and is not designed to specifically measure populations, it is thought to be the largest individual-level database in the country.

The patient register contains the address details of patients; however, depending on the extract, the complete address may not always be available for analysis. Location information derived from just a post code may differ from that derived from the full address where these lie very close to the
boundary of a local authority for example. The quality of the address information is also dependent on individuals keeping their GP up to date with changes, or registering with a new GP if they have made a longer distance move. Any failure or lag in updating address information results in measurement error arising in the location information on the patient register at a particular point in time. For persons who “live” at more than one address the situation is more complex. For example some children at boarding school register with a GP with their boarding school address, others with their parental address.

![Diagram illustrating population and migration statistics](image)

Figure 6.3 Relating the patient register to the usual resident population in area $j$ at time $t$

The coverage of the patient register is all persons registered with a GP in England and Wales. The relationship between the population covered in this data set and the usual resident population is illustrated in Figure 6.3. Here, we see that for a specific area at a specific time, the population captured in the patient register may differ from the usual resident population as a result of the following:

- The inclusion of persons staying in England and Wales for less than 12 months;
- The omission of persons who are not registered with a GP, more prevalent amongst young adult males;
- List inflation (not removing patients when they have moved);
- Erroneous list cleaning (removing patients when they have not moved);
- Issuing duplicate NHS numbers;
- Issuing a new NHS number to someone who is already on the list;
- Lags in re-registration; and
- Lags in international migrants registering.

In terms of timing, the Patient Register extract used by ONS is taken one month after the date at which the usual resident population is estimated to counteract some of the lag in changing GP registrations following moves.

### 6.2.2 Lifetime Labour Market Database (L2)

The National Insurance Recording and PAYE (Pay-As-You-Earn) System (NPS) holds data on all people who have registered or been given a National Insurance Number. The Lifetime Labour
Market database (L2) is a one per cent extract of this data source containing longitudinal information on over 800,000 individuals since 1975. The addresses of individuals in the L2 are obtained from the Customer Information System, which is administered by the Department for Work and Pensions (DWP) and Her Majesty’s Revenue and Customs (HMRC). This data set contains information on most working-age adults who are paying tax, National Insurance or receiving benefits or tax credits, and includes information on age, sex and residence.

The longitudinal nature of the L2 means that it contains a range of address information that can be used to assign a location to individuals. However, the L2 is based on information about a whole tax year. Address start and end dates are supplied and these are used to generate the first address, longest address and last address in each tax year. A decision therefore has to be made about which address to use as the best proxy for address at time \( t \). For example given the proximity of the mid-year point to the beginning of the tax year first address or longest address would probably be the most appropriate candidates.

Like the patient register, the L2 is dependent on individuals notifying changes in their address to DWP or HMRC. This may occur directly by notifying changes in circumstances, for example change of name on marriage, or may happen as a result of them telling their employers who also pass information to DWP or HMRC. So as well as needing to make a decision regarding which is the most appropriate address to use, there may be lags in the latest address being notified. The coverage of the L2 is a one per cent sample of all persons who have ever had a national insurance number (NINo). As NINos are allocated at age 15¾ years, children under this age are not covered.

Figure 6.4: Relating the Lifetime Labour Market database (L2) to the usual resident population in area \( j \) at time \( t \)

As shown in Figure 6.4, the concept of the usual resident population in a specific area at a specific time may differ from the corresponding population captured by the L2 as a result of:

- The lag between immigrants arriving in the UK and registering for a NINo;
- The lag between people moving and HMRC/DWP being advised of that move;
- Non-coverage of children under 15¾;
- Short term immigrants who are registered for a NINo; and
- Persons who have emigrated.
In addition to the coverage issues, timing differences and measurement error arising from lags in the updating of address information, the L2 is also affected by sampling error as it is based on only one per cent of all records on the NPS.

### 6.2.3 English School Census

The English School Census covers all children in state education, representing 92 per cent of school age persons in England. Younger children are also included in the dataset; however, the coverage for this age group is poor as only a small proportion of children attend maintained nurseries. Each pupil has a unique reference number, which means that records for the same pupil can be linked across years from 2002 onwards. For each pupil, geographic referencing using their full address is carried out providing very accurate information about their location. However, lags in updating address information may remain, and moves between the date of the English School Census and the mid-year would not be included.

![Diagram](image-url)

**Figure 6.5: Relating the English School Census to the usual resident population in area \( j \) at time \( t \)**

Figure 6.5 demonstrates how the population covered by the English School Census relates to the usual resident population. The size of the English School Census population is much smaller than the usual resident population because it covers only children. The other main differences arise from:

- Lags in notification of new address details to the school;
- Presence of short-term immigrants on the English School Census;
- Children attending independent schools, home educated or in pupil referral units;
- Children below school age attending private nurseries;
- Children aged 16–18 years who have left school – may vary from area to area depending on education provision; and
- Timing differences – the English School Census is taken in January each year, ONS produces the usual resident population for the mid-year.

### 6.2.4 Higher Education Statistics Agency data

HESA keeps a record of all students attending government funded higher education institutions in the UK by age and sex and provides information on the term time and home addresses. For the
student population, this data set is considered highly reliable. However, it does not cover the population not attending government funded higher education institutions. The usual resident population includes students only at their term time address. The HESA data also includes some information on employment after graduation; however, there are gaps in this dataset that make it less reliable.

The relationships between HESA data and the usual resident population in a particular area and time point are illustrated in Figure 6.5. HESA data are collected for each academic year as a whole, with a reporting period from 1 August in one year to 31 July in the next, however there is no specific date to which the data refer within this period. Data are submitted in October after the academic year to which the data refer.

![Diagram of HESA data and usual resident population relationships](image)

Figure 6.6 Relating HESA data to the usual resident population in area \(j\) at time \(t\)

In comparison to the coverage issues associated with the L2 and the English School Census, HESA data are different in that they do not relate to a specific age group. In practice, very few students are under 18 years of age and the majority are within the 18 to 24 year old age group. However, there are students outside these age groups that need to be accounted for. Note that the issues concerning timing and lags are similar to those for the other sources.

6.2.5 The Annual Population Survey

The Annual Population Survey is comprised of households living at private addresses for five consecutive quarters, meaning that if the household or individuals in that household move, the address is still included in the sample and the new occupants are interviewed (Johnson 2009). These households are divided into five ‘waves’ of approximately 11,000, so in any given quarter, one ‘wave’ receives their first interview, one wave their second interview and so on, the first being a face to face interview and subsequent interviews being held over the phone (ONS 2010a). The Annual Population Survey is useful for tracking the characteristics of the national population over time. However it cannot be used to directly estimate the population, as it requires an independent population estimate to calculate the survey weights. Although the sample size is relatively large, estimates for small sub-sections of the population, including local areas have large levels of sampling error.
6.2.6 Summary of sources in relation to the framework

The NHS patient register provides good coverage of the population and represent the timeliest source of administrative data currently available. Alternative sources such as the National Insurance Recording and PAYE System, the English School Census, Higher Education Statistics Agency data and the Annual Population Survey provide comprehensive data but only for subsections of the population. In combination, all these sources have potential to provide a comprehensive and timely source of information about population change and distributions in the UK. In order to achieve this, we have to use all our knowledge of the coverage and measurement error associated with each of the sources to process them individually and to combine them in a way that moves us closer to the concept of usual residence that we are trying to measure. The coverage, or representation error, within all the sources is a product of a complex web of issues that are difficult to disentangle. For areas where populations are relatively stable these may have a less significant impact than those areas experiencing high levels of population turnover or mixing.

6.3 Estimation methods

6.3.1 Introduction

One of the main statistical series delivered by ONS and other UK National Statistical Offices are the estimates of the usual resident population for the 30th June in each year. This includes the UK and the constituent home countries (England, Wales, Scotland and Northern Ireland), local authorities, electoral areas (e.g., wards), health authorities (Primary Care Trusts, Clinical Commissioning Groups) and small statistical areas (e.g., Lower level Super Output Areas). As illustrated in Figure 6.7, these population estimates are produced directly by a census or indirectly through demographic accounting. Demographic accounting begins with a base population (usually from a recent census) and, over time, adds births, in-migrants and immigrants and subtracts deaths, out-migrants and emigrants (see Figure 6.8 and Appendix 1). The information about internal movements, deaths and births are obtained from administrative registers, while international movements are obtained primarily from a passenger survey.\textsuperscript{12}

Censuses have provided, up to now, the population benchmark at the start of each decade for rolled forward population estimates (Figure 6.8). The population census has also been used to revise population estimates. Before publication of census outputs, validation checks are carried out. One of these is a comparison of census populations, corrected for under- or over-enumeration,\textsuperscript{12}

\textsuperscript{12} Refer to Chapter 8 for more details about ONS method for estimating international migration.
with the rolled forward population estimates from the previous census. In England and Wales in 2001, the checks indicated a census population about 1.2 million lower than was anticipated from the rolled forward population estimates. After a set of local studies which corrected the populations of selected local areas, the census count was adopted as the basis for the mid-2001 population estimate.

Figure 6.8 The demographic accounting steps ONS uses to estimate subnational populations

In terms of the framework presented in Figure 3.1, the main concept is the usual resident population. Ideally, the estimates would utilise information on the stocks and flows of usual residents only, i.e., temporary workers, business travellers and visitors would be excluded. In reality, the estimates are obtained from either a census (i.e., direct count of the population) or from rolled forward estimates using registration information on births, deaths and internal migration and a combination of survey and administrative information on international migration flows.

Not all of the demographic components used to produce the rolled forward estimates are aligned with the usual resident concept of population. The internal migration data (Chapter 7) are based on changes of address notified via GP registrations (implying moves) that occur across administrative boundaries during the time period. The international migration data are aligned with the concept of changes in usual residence but the survey used to obtain this information is only robust at the national level. To obtain subnational estimates of immigration, a new method has been developed for distributing the national level to subnational areas by incorporating information obtained from several administrative sources, depending on the type of migrant. This was used to produce indicative population estimates in 2011\(^1\) and the intention is that it will be adopted formally from 2012 after publication of the 2011 Census. For emigration, there are no subnational administrative sources, so the estimates are obtained by using a propensity to move model that relies on covariate information.

\(^{13}\) For an overview of this methodology see Chapter 8 on international migration. More detail is available from http://www.ons.gov.uk/ons/guide-method/method-quality/imps/improvements-to-local-authority-immigration-estimates/index.html
6.3.2 The inclusion of migration in demographic accounting models

ONS has relied on applying demographic accounting equations to build estimates forward from the most recent census. These equations can take two forms, called movement and transition respectively, depending on how migration flows between areas are measured (Courgeau 1973, Ledent 1980, Rees 1985 and Rees and Willekens 1986). In the movement approach, migration is measured as an event of relocation at a point in time; in the transition approach, migration is measured as a transition between two points in time.

Figure 6.9 illustrates the difference between the two concepts. The vertical axis represents space and is divided in two regions. The horizontal axis represents time (one time interval). The lines on the graph (A, B) plot the location of two people. Person A starts in Region 1 and migrates to Region 2 at time \( t + 0.7 \) and then remains there to be recorded in Region 2 at time \( t + 1 \). Person B starts in Region 2 and migrates to Region 1 at time \( t + 0.2 \) but then migrates back to Region 2 at time \( t + 0.4 \). These two persons make one move from Region 2 to Region 1 and two moves from Region 1 to Region 2. Person A makes one transition, from Region 1 at time \( t \) to Region 2 at time \( t + 1 \). Person B is recorded in Region 1 at time \( t \) and at time \( t + 1 \) and so fails to make a transition. Note that the net migration between regions is +1 for Region 2, whether migration is measured as a move or as transition.\(^{14}\)

Examine Figure 6.9 in isolation it may seem that when estimating the population at time \( t + 1 \) it does not matter whether the movement or transition approach is used to calculate internal migration. Figure 6.10 develops the model further to include births, deaths and immigration. As before, Person A migrates from Region 1 to Region 2, which represents both one move and one transition. Person B migrates from Region 2 to Region 1 and then dies. In the movement framework, the move to Region 1 would be recorded and the death would be recorded in Region 1. In the transition framework, no transition to Region 1 would be recorded and the death would be recorded in Region 1 incorrectly. Person C arrives in Region 1 from outside the UK, moves to

---

\(^{14}\) Rees (1985) explains the relationships between moves and transitions in more detail and also explains a third concept, that of last migrant/last migration, which is widely used in surveys and censuses but which is difficult to use in population estimation.
Region 2 and then back to Region 1. Assuming their arrival in Region 1 was recorded they would represent three moves, but only one transition (i.e., abroad to Region 1). Person D is born in Region 2 then migrates to Region 1. The birth would be recorded in Region 2 and the move to Region 1 recorded in the movement framework. However, in the transition framework there would be no transition recorded (i.e., because there was no place of residence one year previously). This extension highlights the importance of using a consistent approach across the different elements of population change.

![Time-space diagram illustrating different migration measurement concepts including births and deaths](image)

Figure 6.10 A time-space diagram illustrating different migration measurement concepts including births and deaths

In the context of UK population estimation, births and deaths are recorded at the local authority in which they occurred, i.e., the location relevant to the movement framework. As these are the better quality data sources in comparison with the information available to measure migration, it is appropriate to make adjustments to the migration data to ensure they too fit the movement framework, rather than try to estimate births and deaths to fit the transition approach. Chapter 7 explores in more detail how sources of information available to measure migration can be used to produce data related to transitions and/or moves.

6.3.3 Handling special populations: The armed forces, prisoners and school boarders

The population estimation methods described above (and in the Appendix) depend on being able to measure all demographic components of change for all population groups. This is true for births and deaths, where the registration has legal force. This is not true for the internal and international migration components. For instance, the NHS register does not cover the Armed Forces or the Prison population. Flows between the NHS Register and these two sub-populations are recorded but specific origins for migration into the NHS register and specific destinations for migration from the NHS register are not available. So migrations into military bases or between bases are not recorded. A third group who are treated as a special population are school boarders, who are assumed to remain registered with a GP near their home address rather than their school address. However, ONS is aware that this is not always the case, and is investigating this further.
The current approach to estimating populations is to subtract, at the start-of-time interval, the Armed Forces population, the prisoner population and the school boarders from the start population stock and then add fresh stock estimates to the end population. The data for these populations are supplied by the Ministry of Defence, the Ministry of Justice and Department for Education, respectively.

6.3.4 Summary

This section has explored current estimation methods for the usual resident population in the UK. In Section 6.1, we explored how some of the data sources currently available map onto the definition of usual residence. As noted at the beginning of this chapter, ONS through the Migration Statistics Improvement Programme and Beyond 2011 Programme is currently exploring further opportunities for using alternative data sources to quality assure existing methods, or to estimate mid-year populations by using a different approach in the future.
7. **INTERNAL MIGRATION ESTIMATES**

7.1 **Introduction**

In Chapter 6 we reviewed the potential sources and current procedure for population estimates based on the usual residence definition. We also explored the difference between using movement and transitions to measure migration, and explained the need to use moves data in the UK to coincide with data available on births and deaths. This chapter explores the potential sources for measuring internal migration, maps them against both the measurement concept (moves) and the population concept (usual residence) and goes on to explain how the current methodology uses a range of sources to get closer to both of these.

7.2 **Sources of data**

7.2.1 **Introduction**

At present, there is no compulsory system to record movements of the population within the UK. To estimate this information within England and Wales, ONS uses a combination of three administrative data sources in England and Wales; namely the Patient Register Data Service (PRDS), the National Health Service Central Register (NHSCR) and the Higher Education Statistics Agency (HESA). Similar to the estimation of population, there are other potential data sources that can also be used to inform UK internal migration estimates (see Figure 7.1), including, for example, the census, the National Insurance Recording and PAYE System, the English School Census and the Annual Population Survey.

![Figure 7.1 Main sources of data that can be used to inform UK internal migration estimates](image)

---

15 Information from this chapter draws directly from the ONS report ‘Internal Migration Estimates’ published in November 2011 (ONS 2011e).

16 For a general understanding of internal migration statistics, useful sources are Rees and Willekens (1986), Rogerson (1990), Boden et al. (1992), Stillwell (1994), Bell et al. (2002), Fotheringham et al. (2004), Morrison et al. (2004), UK Statistics Authority (2009), Dennett et al. (2007), Stillwell et al. (2010) and Smith et al. (2010).
7.2.2 Patient Register Data Service (PRDS)

PRDS is the main component of the mid-year internal migration estimates. Every Primary Care Trust (PCT) in England and Wales holds information on the patients registered with NHS GPs within their area of responsibility. By obtaining a download from each patient register on an annual basis and by combining all patient register extracts together, it is possible to create a total register for the whole of England and Wales. Comparing NHS numbers in one year with those of the previous year enables identification of people who change their postcode during the year. A migrant is defined as a person who, between one year and the next, changes their area of residence. This final step leads to an estimate of the number of people who have moved during the year, from the number who have changed their postcode following a move. The data are downloaded from the patient registers at each health authority as at the end of July each year to enable migration estimates to be made for the year ending 30th June each year. This is consistent with the assumption that people delay registering with a new GP by approximately a month after they move.

The PRDS captures transitions and, thus, does not record all moves that may have occurred during the year (see illustrations in Section 6.2.2). As highlighted in Figure 7.2, there are several things that one must account for in order to match the PRDS data to all moves required for the population accounting model. For instance, one important group that is missed entirely is the migration of babies born during the year (i.e., those aged less than one year). Other moves which would not be captured by comparing the yearly PRDS files include immigrants, who arrive, register and move within their first year, emigrants who move during the year before they leave and those enlisting in or leaving the armed forces. To overcome the issue concerning multiple moves during the year, ONS uses information from the NHSCR, which records all moves between Health Authorities throughout the year, to supplement the PRDS (see Section 7.2 below). (The PRDS is preferred over the NHSCR as the primary data source in estimating internal migration because it provides a more detailed level of geography).

Figure 7.2 Relating the Patient Register Data Service population transitions to all moves in or out of area \( j \) for a one-year time period
7.2.3 Higher Education Statistics Agency

One of the known limitations of relying on GP registration data is that young people, particularly young men, can be slow to change their registration when they move. An important reason for migration among young people is to attend a course at a higher education establishment, so this limitation of the current internal migration estimation process is a key issue in the estimation of internal migration for this population sub-group. In May 2010, ONS introduced an additional adjustment for students based on HESA data to improve this undercount. HESA data contains records for students registered at higher education establishments and includes both term-time and domicile address variables (see also Section 6.1.4). An adjustment is made to both first year students moving to higher education establishments and moves made by students at the end of their studies.17

7.2.4 Other data sources

The English School Census covers all school children in state education, representing 92 per cent of school age people in England (see also Section 6.1.3). Each pupil has a unique reference number, which means (within the English School Census) one is able to link records for the same pupil, across years from 2002 onwards. For each of the pupils, a full home postcode is recorded, providing a relatively fine geographic detail of their place of usual residence.

The Lifetime Labour Market database (L2; see also Section 6.1.2) is a longitudinal dataset covering 800,000 individuals per tax year since 1975. It is a 1% extract of data from the National Insurance and PAYE System, which holds data on all persons who have registered or been given a National Insurance Number. The addresses of individuals in the L2 are obtained from the Customer Information System (CIS). The CIS is administered by DWP and HMRC.

Migration data are available from the census in the Special Migration Statistics (SMS) tables which give a breakdown of all migrations that occurred in the year before the census, the latest data currently available is internal migrations for 2000/01. The census tables are available by age and sex at regional, local authority and ward level across the UK. Like the PRDS, the census migration data captures transitions (i.e., place of residence on Census night by residence one year previously).

The Annual Population Survey is comprised of households living at private addresses for five consecutive quarters, meaning that if the household or individuals in that household move, the address is still included in the sample and the new occupants are interviewed (Johnson 2009). These households are divided in to five ‘waves’ of approximately 11,000, so in any given quarter, one ‘wave’ will receive their first interview, one wave their second interview and so on with the first being a face to face interview and subsequent interviews being held over the phone (ONS 2010a). Age and sex detail are available in the APS for members of each migrating household that are interviewed, however, the geographic detail is limited because of the relatively small sample size (and that most people do not migrate within a given year).

7.2.5 Summary

An accurate origin-destination-age-sex matrix is the target for statistics on internal migration, and the most comprehensive source for this level of detail is currently the census, albeit for one year out of ten, and for transitions and not moves. For observing annual movements, the combination of PRDS and NHSCR represents the most timely and accurate data currently available. Alternative sources of internal migration data, such as HESA and the English School Census provide comprehensive data but only for subsections of the population. However, when combined with NHS data, they offer opportunities to improve the estimates of internal migration.

17 For more information on the student adjustment methodology please see ONS (2011f).
7.3 Annual mid-year estimation method

7.3.1 Introduction

ONS publish interregional migration estimates on a quarterly basis. These estimates are based solely on NHSCR data. These estimates are timely, published around eight months after the reference period, but are only available at the regional and Former Health Service Authority (FHSA) level. Annual mid-year estimates of internal migration at local authority level are based on a combination of NHSCR, PRDS and HESA data. ONS considers the annual mid-year estimates of internal migration consisting of a combination of administrative data to be more complete, and it is these estimates which are used in the calculation of mid-year population estimates (refer to Chapter 6).

Careful investigations into the feasibility of using PRDS migration data were carried out by Scott and Kilby (1999) and Chappell et al. (2000) which established that patient registers were a sound source of internal migration information. Regular reports on the data sets and guides on how to obtain them are now published in the Internal Migration within England and Wales, year ending June <year> series (e.g. ONS 2009a, ONS 2011a). In terms of outputs, four tables are published on the ONS website each year. They include (1) the numbers of moves to and from each local authority in England and Wales from and to the rest of the UK, (2) an origin-destination table of local authority level moves, (3) an origin-destination table of local authority level moves within each region and to each region by age, and (4) the numbers of moves to and from each local authority in England and Wales from and to England and Wales.

As illustrated in Figure 7.3 and described below, the estimation of internal migration includes three main steps: validation of PRDS, estimation of flows amongst England, Northern Ireland, Scotland and Wales, adjusting the PRDS transition levels to match the NHSCR moves, and adjusting for student flows. Once the internal migration estimates are produced, they are then included into the demographic accounting model to produce annual population estimates of usual residents.

Figure 7.3 The ONS method for estimating annual flows of internal migration by age, sex and local authority in England and Wales

7.3.2 Validating the PRDS

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

Following validation, the patient registers are combined to produce a total patient register for England and Wales. Migrants are then identified by comparing the current year’s patient register with that of the previous year. This is done by matching on the NHS number. By comparing the postcode from one year to the next, any patient that has a change of postcode is identified as a migrant. Migrants are also identified where a patient is only on the patient register for one of the two years. Matches are made against data from the NHSCR, NRS and NISRA to identify cross border moves between the constituent countries of the UK.

Age is calculated as 30th June of the current processing year minus the date of birth. Where the age for a migrant is invalid, it is imputed by matching on migrants with similar values in other variables. If other variables are also missing, age is imputed by matching on variables from migrants identified from incomplete records. If a complete match is not found, conditions are dropped until a match is made solely on the destination or origin FHSA. If a postcode is missing, a migrant may still be identified if the FHSA of registration has changed from one year to the next. However, such migrants need to be allocated down to local authority level. An imputation method has been devised which is similar to that used for imputing missing areas of residence in the International Passenger Survey (Chapter 8). The imputation is carried out by finding all other migrants with the same characteristics, e.g., the same origin and destination FHSA, the same origin local authority and the same age and sex. Once these migrants are identified, the missing characteristics in the incomplete record are imputed using a random allocation of local authority from those migrants identified as being similar.

7.3.3 Flows between England, Northern Ireland, Scotland and Wales

Data from the patient registers only enables identification of persons who have moved within England, within Wales and between England and Wales. However, the NHSCR, and data from NRS and NISRA can be used to provide information on moves from Scotland and Northern Ireland. Migrants moving into England and Wales from Scotland and Northern Ireland are identified when the NHS number on the current year’s patient register is not found on the previous year’s patient register and the previous health authority on the current year’s patient register indicates Scotland or Northern Ireland. These migrants are identified and then constrained to data from the NHSCR on moves into England and Wales from Scotland and Northern Ireland.

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

7.3.4 Constraining the PRDS to NHSCR

The migration data based on the PRDS are constrained to the NHSCR estimates at the FHSA level to account for migrants that may be missing. The NHSCR holds information on moves of

---

Reclassifications of postcodes made by the Royal Mail are taken into account to avoid creating false migrants.
migrants aged one and over who have moved between FHSAs and this is used to constrain the PRDS for migrants aged one and over who move between FHSAs. Information on moves of migrants within health authorities is not available on the NHSCR data, which only records moves between FHSAs, and so it is assumed that the volume of scaling up moves between FHSAs is the same as that needed for within area moves.

For moves between FHSAs for migrants aged one and over, the number of moves for each combination of FHSAs in the patient register is constrained to the number of moves for the same combination of FHSAs in the NHSCR by applying scaling factors. To adjust the patient register transitions data for moves within a FHSA, the scaling factors used to constrain between area moves have been used. For each combination of FHSAs, the scaling factor for moves into that FHSA are combined with the scaling factor for moves out of that FHSA to obtain an average scaling factor for moves within that FHSA. This average scaling factor is then applied to all transitions with each FHSA.

Comparing patient registers from two years does not capture any information on migrants aged under one since they do not appear on the patient register in the first year. While the NHSCR data does hold information on moves under one, this information is at FHSA level and the patient register data is at a lower geographic level. However, research has shown that the distribution of male and female migrants aged under one by FHSA in the NHSCR data is strongly correlated ($r^2 = 0.97$) with the distribution of male and female migrants aged one by FHSA in the patient register data. Thus, the distribution of migrants aged one in the patient register data by sex, origin FHSA and origin local authority is used to distribute data from the NHSCR on migrant babies under one by sex, origin FHSA and origin local authority level. Likewise, the distribution of migrant babies aged one in the patient register data by sex, destination FHSA and destination local authority is used to distribute data from the NHSCR on migrant babies under one by sex, destination FHSA and destination local authority level.

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

### 7.3.5 HESA adjustment

In 2010, ONS introduced an adjustment for students calculated from data from the Higher Education Statistics Agency (HESA). This adjustment was designed to better capture the movement of students to and from places of study. The adjustment is applied to the constrained PRDS tables and consists of three adjustments. First, an adjustment is made to the number of moves made by first year undergraduate students based on HESA data on students’ domicile and term-time address. Second, an adjustment is made to capture the number of moves made by students at the end of their studies. The approach uses a range of sources in the absence of HESA data on students’ address in the year after their studies. Finally, an adjustment is made to avoid counting moves of students/former students twice when they do eventually re-register. In the initial implementation an adjustment was made to the number of moves made by first year undergraduate students between mid-2001 and mid-2007 based on HESA data. Additional assumptions were made in the absence of HESA data on students’ term-time address prior to academic year 2007/08. Further detail of this methodology can be found in ONS (2011f).

### 7.3.6 Approaches in Scotland and Northern Ireland

Scotland uses health registers to track internal population movements via the Scottish National Health Service Central Register (SNHSCR) and the Central Health Index (CHI). An inter-council
area matrix of flows is produced by National Records of Scotland (NRS) by using the SNHSCR to record movements of migrants between Health Board areas, combined with the Scotland Community Health Index, which contains the postcode of people registered with an NHS doctor in Scotland, allowing for internal migration to be estimated at council area scale and below. As in England and Wales, the data derived from the SNHSCR are available quarterly but the data do not contain any address detail below the Health Board level. For this reason, the annually downloaded Scotland Community Health Index estimates are controlled to the SNHSCR totals by origin, destination, age and sex (NRS 2010). NRS also match CHI birth records to obtain mother’s postcode at child’s birth in order to account for migration of those under one year of age. No student adjustment is made for inter-council area flows in Scotland. NRS are investigating further improvements to the methodology for Scotland.

Northern Ireland also uses registers to record internal migration. Northern Ireland Statistics and Research Agency (NISRA) estimates flows at Local Government District and Parliamentary Constituency levels using the Northern Ireland Central Health Index (NICHI) which records changes in address when re-registering with a GP. In addition, a student adjustment is made, informed by administrative data sources, by removing people of student age from most Local Government Districts and adding them to a small number of Local Government Districts with centres of tertiary education (NISRA 2007, p.3).
8. INTERNATIONAL MIGRATION ESTIMATES

8.1 Introduction

In Chapter 7 we reviewed the potential sources and current estimation process for internal migration and linked it with Chapter 6 on population estimates based on the usual residence definition. This chapter focuses on the potential sources for measuring long-term international migration, where the moves represent cross UK border changes in usual residence. We then describe the methods for producing national level estimates of immigration and emigration, and their allocation to local authorities.

8.2 Sources of data

8.2.1 Introduction

Countries tend to gather migration data according to their own needs (often for legal purposes) or to be consistent with historical collection methods. Until very recently in the European Union, there have been no real incentives for countries to adjust their data collection methods to provide internationally comparable migration statistics. This means that, in order to understand or predict how international migration between countries evolves over time, one must have a good sense of the various migration data typologies and the determinants of migration. In this section, we summarise the main issues concerning the reported flows in Europe and then discuss the UK situation.

The availability of statistics on international migration flows is conditioned by the existence of a data collection system that has the potential of yielding meaningful statistical information on changes in the country of usual residence. The major types of data sources used to produce statistics on international migration flows can be summarized as follows:

1) population registration systems, including centralised population registers and local population registers;

2) other administrative registers related to foreigners, alien registers, residence permit databases or asylum seeker databases;

3) statistical forms filled in for all changes of residence; and

4) border crossing data collection and other sample surveys.

Some information on international migration flows can also be derived from population censuses, but this source has a number of well-known limitations. The main ones are that they are carried out at longer intervals, e.g., every ten years, and capable of only identifying immigrants, as emigrants are no longer present to be counted in the country of origin. Because of these reasons, migration flow data obtained from censuses are usually not considered for the reporting of international migration flows. They are, however, used to measure, for example, the proportion of the population born outside the UK.

Even if flow data are available, their poor quality may render them inadequate for use. There are two main factors that make international migration statistics unreliable. The first is the under-registration of migrations, which applies in particular to countries where data-collection systems rely on self-declarations of international movements. The second relates to data coverage in which

the data collection system used in a country may not cover the correct target population and so some subsets of migrants are excluded from the statistics (e.g., asylum seekers, students or armed forces) or included (short term migrants). In addition to the above two factors, data may be unreliable if a lot of errors arise during data processing.

In the UK, there is no single source of data that captures all long-term international migration to and from the country. As a result, ONS uses a combination of data from different sources. Each source of data has different characteristics that can be used to help estimate international migration. However, it is important to note that none of the data sources used are designed specifically with this aim in mind. As shown in Figure 8.1, the current estimates of Long Term International Migration (LTIM) are comprised primarily of data from the International Passenger Survey (IPS), Northern Ireland Statistics and Research Agency (NISRA) for flows between Northern Ireland and the rest of the world outside the UK and the Home Office for asylum seekers and refugees. In combination, these sources provide the national level estimates of immigration and emigration (LTIM) by age, sex and other broad characteristics, such as flows by British and non-British citizens. To obtain subnational estimates of immigration, ONS' proposed method (ONS, 2011g) first uses the LFS to produce data for Scotland and England and Wales separately and then utilises other data sources, such as the census, the Higher Education Statistics Agency (HESA) data, the National Insurance Recording and PAYE System and the GP patient register to disaggregate further within England and Wales. Subnational estimates of emigration are more difficult to obtain as there are very few sources of data on these migrants, and none of these are particularly reliable. As a result ONS uses a propensity to move model that relies on covariate information.

As described below, the LTIM estimates of immigration and emigration are used as benchmarks for estimating tables required by Eurostat (see Section 8.3.2) and subnational flows (see Sections 8.4 and 8.5) for informing post-census population estimates of local authorities. Refer to Raymer et al. (2011b) and Boden and Rees (2010a, b), respectively, for more details and methodological
issues concerning the estimation of these patterns. Note that statistics on births, deaths and internal migration are all collected directly at local authority level. International migration flows, on the other hand, are drawn largely from a national passenger survey not designed for high levels of disaggregation. Thus, subnational allocations of the LTIM flows have to be estimated by using auxiliary and covariate information.

Finally, for the international migration component, the destination within the UK for an immigrant, the origin within the UK for an emigrant, age and sex represent the essential variables that feed through to the cohort component model of population estimation. The origin country for an immigrant and the destination country for an emigrant are desirable but are not required to produce population estimates at national or local level. The main emphasis is thus concerned with counting the total number of persons migrating in and out of the UK.

8.2.2 International Passenger Survey

The International Passenger Survey is the main source of information on international migration. It is the only source that captures all migrants in both directions (except asylum seekers, refugees and flows from and to Northern Ireland). Furthermore, the measurement of international migration is aligned with both the United Nations (1998) and Eurostat’s recommended definition for an international long-term migrant, which is “someone who changes his or her country of usual residence for a period of at least a year, so that the country of destination effectively becomes the country of usual residence.”

The international migration flows derived from the IPS are based on a sample of travellers interviewed on entering or leaving the UK via the major air, sea and Channel Tunnel embarkation points. Of the 800,000 interviews conducted in 2010, 0.6% were migrants, giving a sample size of around 5,000 (ONS 2012g). Because the IPS is an intentions-based survey, asking migrants where they intend to go and how long they intend to stay in the country, an adjustment is made for ‘migrant switchers’ and for ‘visitor switchers’. The migrant switcher adjustment takes into account people who state an intention to stay for over 12 months but who stay, in fact, for less than 12 months. The visitor switcher adjustment accounts for people whose intention is to stay for less than 12 months but who remain for over 12 months (ONS 2012h). Switcher numbers are estimated using the response to questions on the IPS of previous migrants and asks “when you last arrived in (left) the UK, how long did you intend to stay (away) for?” (ONS 2012h). In practice, migrant and visitor switchers have tended to be in balance apart from in 2005 and 2006 and this estimation has not altered the total international migration flow greatly.

In Figure 8.2, we relate the migration flows captured by the IPS with all migrants arriving or leaving the UK that are required for population estimation and international comparisons. These numbers, for each year, represent the total number of persons in the UK changing their country of residence. The definition of migration used in the IPS is thus aligned well with the concept of migration required for population estimation at the national level. The main concern is the sample size and the fact that most passengers are not long-term migrants. Migration estimates are generally regarded as unreliable if the standard error is greater than 25 per cent of the estimate. However, these standard errors have some limitations as they are designed for the IPS as a whole, and not specifically for the subset of IPS migrants. Because the IPS data come from a survey, disaggregation by age, sex, country of origin and year, for example, are likely to require smoothing or inferring with auxiliary information.

The flows of asylum seekers and refugees, not captured by the IPS, are not problematic since the Home Office is responsible for tracking these types of migrants and is much more likely to provide accurate information on these types of flows. Likewise, it is believed that NISRA is better able to measure the movement of international migrants across its borders than the IPS, particularly for those originating from (or departing to) the Republic of Ireland.
8.2.3 Northern Ireland Statistics and Research Agency (international migration flows from / to Northern Ireland)

The IPS was used to estimate migration to and from Northern Ireland until 2007. However, there were concerns about the reliability of these estimates. Therefore, from 2008 onwards, ONS incorporated Northern Ireland’s Statistics and Research Agency’s (NISRA) estimations of long term international migration based on their patient registers into the LTIM estimates. At the same time, ONS switched from using data from the Central Statistical Office (CSO) in the Republic of Ireland for estimating flows between the Republic of Ireland and the UK, to using the IPS to estimate flows between the Republic of Ireland and Great Britain. Flows between the Republic of Ireland and Northern Ireland are included in the NISRA estimates. Finally, to prevent double counting, the flows to Northern Ireland reported in the IPS have been removed.

8.2.4 Home Office (asylum seekers and refugees)

The Home Office produces quarterly statistics on border control and asylum (see, e.g., Home Office 2011). The border control statistics include entry clearance visas and grants of an extension of leave to remain. The Home Office also, until recently, administered the Workers Registration Scheme for new EU accession migrants. The statistics on asylum seekers concern applications, support applications, decisions (including dependents), appeals, current numbers being supported, removals and voluntary departures, and acquisitions of British citizenship.

The asylum seekers and refugee data are currently used by ONS to supplement the International Passenger Survey data in the construction of the Long Term International Migration (LTIM) statistics. While the data are considered high quality, they “do not correspond directly to the standard ONS definition of a long-term international migrant” and as such, broad assumptions are
made by ONS about the proportion of asylum seeker applicants that actually correspond to the ONS definition (ONS 2011g, p.2). Under the new methodology for distribution of immigrants to local authorities, asylum seekers and refugees are distributed directly based on the Home Office data.

8.2.5 National Health Service patient register for England and Wales (‘Flag 4’ patients)

In the National Health Service Central Register, an immigrant from outside the UK is identified when a new NHS number is issued for a person previously living abroad. This code is known by ONS as a ‘flag 4’ identifier. Some information is held on country of birth. However, this only includes the four categories of UK, Ireland, Elsewhere and Unknown. The registration of new patients coming from abroad has several issues in relation to the UN definition of international migration and usual residence, such as lags in registration and unknown lengths of time spent in the UK. Refer to Sections 6.2.1 and 7.2.2 for more information concerning patient register data.

‘Flag 4’ statistics feature in the new local authority distribution method recently introduced by ONS for areas in England and Wales in 2011. They are used to allocate ‘other’ migrants to the local authority level.

8.2.6 National Insurance Recording and PAYE System

The National Insurance Recording and PAYE System contains several datasets that are of interest to measuring international migration, particularly flows of immigration. The Migrant Workers Scan contains information on all overseas nationals who have registered and been allocated a National Insurance Number (NINo) from 2002. The database of NINos may be used to augment the estimation of immigration to local authorities in the UK. A NINo is required by overseas nationals intending to work or claim benefits or tax credits in the UK. This includes self-employed persons and students working part time (DWP 2010). NINos are available on some other administrative sources, notably within DWP and HMRC systems.

The Migrant Workers Scan provides counts of overseas nationals with new NINo registrations in each local authority and is useful for identifying immigrants by country of citizenship. Along with the Lifetime Labour Market Database (L2; see section 6.2.2), this database is used to distribute work-related immigrants to local authorities in England and Wales. The concerns with the Migrant Workers Scan are

- Captures only a subset of all overseas nationals;
- Contains lags in registration;
- Arrival dates, addresses and changes in addresses are self reported and not compulsory;
- Includes both short- and long-term migrants together;
- Emigrants are not known; and
- National Insurance Numbers are not available on other administrative sources.

The benefits of using this administrative data are the ability to track individuals over time, including changes in address, and the high level of population coverage. Thus, the MWS provides a rich source of information, both longitudinally and as a snapshot. Furthermore, by using the L2 database, ONS is able to distinguish between short-term and long-term migrants based on their activity over time. This has allowed for a marked improvement in the distribution of migrant workers across local authorities.

8.2.7 Higher Education Statistics Agency (foreign student statistics)

Similar to the L2, HESA data (Section 6.2.4) can be used to identify a specific sub-group of the IPS immigrant population, i.e., persons travelling to the UK in order to study at a publicly funded higher education institution. HESA data provide a useful means of identifying international student migration, as it contains the term time addresses of students, institutions of study, citizenships and a number of other useful characteristics, such as age and sex. HESA data can also be used to
identify the durations of study or stay in the UK. For these reasons, it has been adopted by ONS in the new distribution methodology to allocate students, identified in the IPS, to local authorities in England and Wales.

8.2.8 2001 Census

The Special Migration Statistics (SMS) from the 2001 Census provide a set of immigration flows for 2000/01. The counts are available for immigration taking place in the year before the 2001 Census at the national level and at subnational geographies (Government Office Region, LA and Ward) level. These flows can be disaggregated by age and sex and a similar matrix will be available for 2010/11 from the 2011 Census in due course. The 2001 Census is used to allocate UK born returning migrants to local authorities in England and Wales in the new ONS distribution methodology.

8.2.9 Summary

There are many possible sources of information on long-term international migration in the UK. However, none of them are sufficient on their own. As described in the next section, national level estimates of immigration and emigration are obtained by combining the IPS, adjusted for visitor and migrant switchers, with data from NISRA on flows to and from Northern Ireland and data from the Home Office on flows of asylum seekers and refugees. In Sections 8.4 and 8.5, the methodologies for distributing these national level estimates to local authorities are described for immigration and emigration, respectively.

8.3 National level immigration and emigration

In this section, we first describe ONS’s current method for producing international migration statistics at the national level. We then outline Eurostat’s mandatory requirements for immigration and emigration and discuss some of the issues for meeting these requirements.

8.3.1 Estimation method for estimating long-term international migration flows

As there is no single, all-inclusive system for measuring long-term international migration in the UK, ONS combines information from different data sources to produce estimates of immigration and emigration. The Long-Term International Migration (LTIM) estimates are based primarily on a subset of the International Passenger Survey (IPS), namely those international passengers sampled by the IPS who are migrants entering or leaving the UK by principal air, sea and tunnel routes. The IPS component is supplemented with:

- Home Office administrative data, which is used to calculate an adjustment for asylum seekers and their dependants not counted by the IPS (note, enforced removals are not included in the ONS calculations of international migration);
- An adjustment for visitor switchers (i.e. those who intend to enter, or leave, the UK for less than 12 months but will actually stay, or stay away, for longer) and migrant switchers (i.e. those who intend to enter, or leave, the UK for at least 12 months without those intentions being realised);
- Quarterly National Household Survey data on flows to and from the Republic of Ireland provided by the Irish Central Statistics Office (1991 to 2007); and
- Northern Ireland Statistics and Research Agency data on migration to and from Northern Ireland (from 2008 onwards).

8.3.2 Eurostat’s requirements

This section draws directly from ONS (2011g).
The following information is taken from Article 3 of the European Parliament Regulation (EC) No. 862/2007 (European Commission 2007). Member states are required to supply the following international migration flow data to Eurostat:

- Immigrants disaggregated by (i) groups of citizenship by age and sex, (ii) groups of country of birth by age and sex, and (iii) groups of country of previous usual residence by age and sex;
- Emigrants disaggregated by (i) groups of citizenship, (ii) age, (iii) sex, and (iv) groups of countries of next usual residence.

Country groups are defined as follows:

- EU27 – 27 member states of the European Union;
- EFTA – The European Free Trade Association;
- CC3_07 – European Union Candidate Countries;
- HDC – Non-EU Highly Developed Country;
- MDC – Non-EU Medium Developed Country; and
- LDC – Non-EU Low Developed Country.

In addition to these requirements, member countries are encouraged to supply other migration data, such as immigration flows by country of previous residence, on a voluntary basis.

The IPS, which represents the primary source of the LTIM estimates, is not designed for producing statistics at the level of Eurostat’s requirements. The sample size of the IPS and the number of migrants it captures are too small. Thus, to meet these requirements, ONS must use further estimation methods, which they are currently in the process of developing. Refer to Raymer et al. (2011b) for methodological options concerning the smoothing, repairing or augmenting IPS data.

8.4 Immigration to local authorities in England and Wales

Until June 2010, the ONS method for estimating immigration at the local authority level used the Labour Force Survey (LFS) to allocate gross national IPS flows to Wales and the English regions. LFS statistics on ‘long-term’ migrants calibrated the proportional distribution of flows to each Region and IPS data, smoothed over an extended time-series, was used to allocate immigration flows to intermediate geographies outside London, with the LFS sample size believed to be sufficiently robust to enable it to be used for estimation for the London areas. The final stage of immigration estimation originally involved the proportional allocation of flows to local authority areas using the migrant distributions evident from 2001 Census data. However, this was replaced by a model based methodology using administrative data sources to provide covariate information. The emigration estimation process had a similar hierarchical structure but did not have the luxury of additional data from either the LFS or the census. As an alternative, it incorporated a ‘migration propensity’ model to estimate the distribution of flows at a local authority level.

Based on the ideas and findings from the New Migrant Databank (Boden and Rees 2010a, b), ONS have recently developed a new method to estimate immigration to local authorities based on combining three administrative datasets with the IPS: GP registration of foreign nationals, National Insurance Number registrations of foreign nationals and the Higher Education Statistical Agency’s count of international student numbers (see Figure 8.1). The new methodology removes the requirement for a regional and intermediate geographical hierarchy and maximises the use of ‘local’ administrative evidence on international migrant populations. The use of the three administrative data sources enables the identification of the geographical distributions of different migrant streams (i.e., workers, students and other migrants). An overview of the new methodology for distributing immigration to local authorities in England and Wales is presented in Figure 8.3.

---

21 This section draws directly from ONS (2011g).
Figure 8.3 Overview of method for estimating immigration flows at local authority level

The main features of the new distributional methodology are as follows (Swier 2011):

- The key principle is to achieve the closest possible mapping between the IPS and the available administrative data.
- The local authority estimates are based on distributions and not the actual administrative counts. Thus, the total population estimate for England and Wales does not change.
- 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.
- Record linkage is used both within and between sources to minimise definitional differences and double counting.

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) 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 is used to distribute asylum seeker flows identified in the LTIM.
• Flag 4 data from the GP Patient Register Database is used to distribute the remaining migrants.

The new method is being applied for mid-2005/06 to mid-2009/10 years to provide a sufficient set of trend data for the 2010-based Sub-National Population Projections and because some of the administrative data are not available for earlier years.

8.5 Emigration from local authorities in England and Wales

As mentioned previously, emigration is the demographic component with the least amount of information available. To estimate emigration from local authorities, the IPS and migrant switcher estimates are first disaggregated into an intermediate geography level based on IPS data averaged over three years (ONS 2010b). These estimates of emigration are then distributed to local authorities according to the estimated coefficients of a Poisson regression model fitted to the averaged IPS data with a consistent set of predictor variables used over time. An assumption is made that the estimated coefficients at the intermediate geography level are applicable at the local authority level. Visitor switchers are processed separately, and asylum seekers distributed according to data from the National Asylum Support Service (NASS).
9. POPULATION AND MIGRATION ATTRIBUTES

9.1 Understanding society

Until now, the report has focused on the production, issues and needs for the mid-year estimates produced by ONS. These include population totals by age, sex and local authority. To understand population change and differences across geographies in the UK, however, we need more detailed information about the populations. In this chapter, we outline the needs and sources for more detailed information about populations in the UK. In general, we can think about population change being influenced by social and cultural contexts, economic contexts and environmental contexts and the intersections between them as outlined in Figure 1.2.

Societies and their demographic structures are in constant flux, resulting in specific social challenges over time. For example, people are more mobile than they have ever been, with advances in transport and communications. In 2004, there was a significant expansion of the European Union, with the accession of eight new member states. As a result, there has been a significant inflow of migrants to the UK from both within and outside the European Union. We are also living longer, with advances in health and with more information available to people about leading healthy lifestyles. However, running alongside this, there are substantial variations found in health conditions and life expectancies within the UK by geography and socio-economic characteristics (Davey-Smith et al. 2002). Explaining and understanding these social changes is a major challenge and research interest for academics and policy makers alike.

In the commercial sector, companies use detailed information about populations, such as age, housing and purchases, to target their goods and services at specific groups of people (Boyle and Dorling 2004). In public policy, population statistics and covariates are used to monitor the impact of policy change and to determine the success and impact of government interventions and policies.

There has been a rise in levels of ethnic diversity in the UK in recent decades, with important regional and local variations in the ethnic composition and rate of change of the sizes of various ethnic groups (Rees and Butt 2004). Academics and policy makers are interested in how ethnic groups differ in terms of socio-economic conditions and across generations.

The above represent just a few examples of the demographic and social context for population statistics, all of which have important social policy implications and create research questions that need to be addressed. For these changes to be understood and researched, it is important to have statistics on the size, location (outlined previously) and characteristics of the population in the UK. The following sections outline the attribute data within the census and large social surveys in the UK.

9.2 Key attributes and sources of data

9.2.1 Introduction

The key attributes of the population can be considered under three headings. First, the social and cultural attributes of the population relate to variables that identify the key characteristics of individuals, such as nationality and marital status. Second, economic attributes relate to the individual and define the elements of their life associated with work and economic activity, such as the industry-sector they are employed in and the number of hours they work every week. Finally, environmental attributes can be used to group household and family groups into the types of areas they live in and are usually made up of variables derived from the attributes of the individuals in that group or relate to the type of accommodation the household lives in. As shown in Figure 9.1, there is some cross over between the groups, for example, the employment attributes for the respondent to a survey are often used as a proxy for the economic status of the family or
household. Likewise, social attributes are often assigned to families and households based on the response of an individual who is surveyed on behalf of the group.

### Key Variables Collected from Census and Surveys

<table>
<thead>
<tr>
<th>SOCIAL</th>
<th>HOUSEHOLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Age</td>
<td>• Total number of people in household</td>
</tr>
<tr>
<td>• Date of birth</td>
<td>• Age of people in household</td>
</tr>
<tr>
<td>• Sex</td>
<td>• Total number of family units in household</td>
</tr>
<tr>
<td>• Marital status</td>
<td>• Number and age of children</td>
</tr>
<tr>
<td>• Nationality</td>
<td>• Composition of household (single parent, cohabiting couple etc)</td>
</tr>
<tr>
<td>• Country of birth</td>
<td>• Household economic activity</td>
</tr>
<tr>
<td>• Year arrived in UK</td>
<td>• Tenure</td>
</tr>
<tr>
<td>• National identity</td>
<td>• Furnished/ non-furnished</td>
</tr>
<tr>
<td>• Ethnicity</td>
<td>• Number of rooms</td>
</tr>
<tr>
<td>• Religion</td>
<td>• Number of bedrooms</td>
</tr>
<tr>
<td>• Sexual identity</td>
<td>• Type of accommodation (house, flat, communal establishment)</td>
</tr>
<tr>
<td>• Wellbeing</td>
<td>• Amenities (own bath/shower, central heating)</td>
</tr>
<tr>
<td>• Health</td>
<td>• Lowest floor level (basement, 1st floor etc)</td>
</tr>
<tr>
<td>• LTI</td>
<td>• Type of communal establishment</td>
</tr>
<tr>
<td>• Provision of unpaid care</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EMPLOYMENT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Economic activity</td>
<td>Active = full time, part time, self employed, unemployed, full time student; Inactive = retired, student, looking after home, LTTI, disabled</td>
</tr>
<tr>
<td>• Industry sector</td>
<td></td>
</tr>
<tr>
<td>• Occupational group</td>
<td></td>
</tr>
<tr>
<td>• NS-SEC class</td>
<td></td>
</tr>
<tr>
<td>• Hours worked</td>
<td></td>
</tr>
<tr>
<td>• Travel to work time</td>
<td></td>
</tr>
</tbody>
</table>

Figure 9.1 Summary of key variables collected from censuses and surveys

In the UK, the most comprehensive source of attribute information is the decennial census which aims to cover 100% of the population. In the intervening periods between censuses, surveys provide more timely information based on responses from a sample of the population, albeit with limited geographical detail. There are also many surveys carried out in the UK, as described in Section 9.2.3 below, including two panel surveys. Surveys in the UK provide a huge amount of social, labour and household attribute information about the population over time.

#### 9.2.2 Census

Censuses are viewed as the most important benchmark for providing attribute data on the UK population. They contain nearly 100% coverage of the population and a large number of social, labour force and family/household variables. Information from censuses can also be compared over time, provided the variables do not change much, to track changes in the population attributes. For example, one could examine how average household size has changed according to the type of the housing in each of the local authorities in the UK from one decade to the next.
9.2.3 Surveys

ONS have produced a definitive list of surveys carried out in the UK (ONS 2012e) covering a wide variety of themes and a similar listing has been compiled by ESDS Government of all surveys they support (ESDS Government (2012).

A notable recent development in the collection of social, labour and household attributes of the population comes in the form of the Integrated Household Survey. This survey brings together a number of key questions asked across a range of social surveys conducted by ONS, and purports to have achieved a sample size of 450,000 individuals between April 2009 and March 2010 (ONS 2010c). A set of ‘core’ questions are asked in the individual surveys and these are deposited on the Integrated Household Survey, while a number of ‘bolt on’ questions, which are not included in the Integrated Household Survey, are reported in the individual surveys. ONS (2010a, p.2) report that “the aim of the Integrated Household Survey is to produce high-level estimates for particular themes to a higher precision and lower geographic level than current ONS social surveys.” The Integrated Household Survey began in January 2008 and during its development the ‘core’ questions have been asked as part of the General Lifestyle Survey, Living Cost and Food Survey, Opinions Survey, Labour Force Survey (LFS) / Annual Population Survey (APS), Life Opportunities Survey and the English Housing Survey. Following the first reported results for 2009/10, some surveys containing ‘core’ questions were dropped from the Integrated Household Survey, so from April 2011 the Integrated Household Survey was made up of core questions asked on three surveys: the General Lifestyle Survey, Living Cost and Food Survey and the LFS / APS. It has been reported that the General Lifestyle Survey will be phased out from autumn 2012 and “it is expected that the composition of the Integrated Household Survey will be flexible with some surveys leaving the Integrated Household Survey and others entering each year” (Walthery 2011, p.3).

The Integrated Household Survey is made up of a core of around 100 questions. The majority of these questions are present on all component surveys, with a small number either being unique to or omitted from a particular survey, or worded differently between surveys. The number of attributes collected as part of the Integrated Household Survey is largely comparable to the census, although it must be remembered that the response is based on a much smaller sample of the population. Each of the surveys have their own methodology and sampling design meaning that both clustered and un-clustered sampling is used (ONS 2010c, p. 4).

The General Lifestyle Survey (known as the General Household Survey until 2008) covers Great Britain. It is made up of questions relating to the household, completed by the household reference person and individual questionnaires, completed by all people in the household aged 16 and over. In 2009/10, the sample size for the General Lifestyle Survey was 18,033 people (ONS 2010d), including students living in halls of residence, who were identified as part of the household being interviewed (Dunstan 2009). However, with regard to coverage, a limitation of the General Lifestyle Survey is that it does not include communal establishments as part of the survey. Members of the General Lifestyle Survey are interviewed for four consecutive years and individuals are traced to their new household if they move (ONS 2010d). In addition to the core variables reported in the Integrated Household Survey, the General Lifestyle Survey gives an insight into personal attributes not found on other surveys, such as the amount of alcohol that people drink, the use of medical services and whether or not they pay in to a pension fund. The General Lifestyle Survey also reports on household variables related to deprivation, such as access to central heating and a measure of overcrowding through an occupancy rating.

The Living Cost and Food Survey is a household survey. Each individual in the household, aged 16 years and older, is asked to keep a diary of daily expenditure for two weeks, coupled with questions on regular expenditure, historic large and infrequent purchases, income and personal demographic data (ONS 2010e, p. 17). Between 1957 and 2001 two surveys, the Family Expenditure Survey and National Food Survey captured data on expenditure. These were combined in 2001 to form the Expenditure and Food Survey before finally becoming the Living
Cost and Food Survey in 2008 (ONS 2010e). The sample size of the Living Cost and Food Survey for 2009/10 was 11,989 people living in 5,241 households (ONS 2010e).

The APS is a household level dataset which comprises the quarterly LFS and the APS dataset. The primary purpose of the LFS is to “produce person-level statistics (such as employment, unemployment and economic inactivity levels and rates) broken down by personal characteristics (such as age, sex and ethnicity)” (Johnson 2009, p.2). It also produces family and household level statistics that combine the economic activity status of family/household members.

The APS dataset compiles a set of sample-boosts for the LFS, carried out in England, Scotland and Wales known as the Local Labour Force Survey (LLFS). LLFSs are based on the LFS structure and are produced for England (ELLFS), Scotland (SLLFS) and Wales (WLLFS) in order to provide more accurate labour force statistics at the sub-regional level (Werner, 2006). Each household in the LLFS is interviewed annually for four years (ONS 2008). By combining individuals in waves one and five from four consecutive quarters of the LFS and data from the annual LLFS the APS dataset is created (Ashton and Kent 2008).

In 2009/10, the combined sample size of the LFS and APS dataset was 334,206 individuals, representing a large proportion (74 per cent) of the Integrated Household Survey dataset (ONS 2010c). The individuals are drawn from approximately 213,000 households (53,000 in the LFS sample, 160,000 in the APS sample) (Johnson, 2009). As both the APS and LFS are aimed at private households, the coverage excludes many groups living in communal establishments, “except those in NHS housing and students in halls of residence, where inclusion takes place at their parents’ address. Members of the armed forces are included if they live in private accommodation” (Ashton and Kent 2008, p.1). The implication of omitting communal establishments from the LFS was an estimated undercount of 89,600 people in 2009/10 (ONS 2010c, p.13).

The LFS sample is comprised of households living at private addresses for five consecutive quarters, meaning that if the household or individuals in that household move, the address is still included in the sample and the new occupants are interviewed (Johnson 2009). These households are divided into five ‘waves’ of approximately 11,000, so in any given quarter, one ‘wave’ will receive their first interview, one wave their second interview and so on. The first interview is face to face with subsequent interviews being held over the phone (ONS 2010c). There is a Eurostat requirement for member states to collect information on their labour force and the LFS fulfils the UK’s EUROSTAT requirement (ONS 2009b) and the required variables are produced on a quarterly basis.

Individual and family/household datasets are produced for both the LFS and APS. The household datasets include all variables found on the person level datasets (except earnings variables) as well as a number of derived variables to allow for the combination of economic activity in the family/household (ONS 2008). As well as the core attributes found in the Integrated Household Survey, the LFS / APS dataset contains labour force variables relating to income, benefits, hours worked and sickness. Only the LFS portion details income from employment. In 2008, the EU required ONS to implement a special migration module as part of the LFS (ESDS 2008).

The Life Opportunities Survey addresses the social barriers to participation that disabled people in Great Britain experience and is designed to compare the experiences of disabled people with those of non-disabled people (Howe 2010). It is a longitudinal study comprised of a sample of 23,368 people in 2009/10, drawn from 37,500 households (DWP 2011a). The target demographic of the survey is persons aged 16 years and over, however, demographic attributes of children in the household are also collected (DWP 2011a). In addition to questions relating to access, it is the only survey that asks detailed health questions of respondents and asks about provision of unpaid care and the household’s ability to cope financially.
The English Housing Survey is commissioned by the Department for Communities and Local Government and is split into data on housing stock and data on households. The latter provides detail on tenure patterns, overcrowding and under occupation for households in England (ESDS 2012). The English Housing Survey had a sample size of 40,753 people in 2009/10 from 32,100 addresses (ONS 2010f). It contains questions that relate to measuring deprivation, such as whether a household has the sole use of a bathroom, shower and toilet and if the accommodation is shared or not. Household income and car ownership variables are also reported.

The ONS Opinions Survey (formerly the ONS Omnibus Survey) is carried out monthly and is a multipurpose survey to obtain information on a variety of topics too brief to warrant surveys of their own. The 2009/10 sample surveyed 20,981 individuals from 2010 addresses per month (ONS 2010g).

The Family Resources Survey, run by the Department for Work and Pensions (DWP) provides detailed information on income and employment in the UK (DWP 2011b). The sample size in 2009/10 was 75,279 people contained within 25,205 households (DWP 2011b). In addition to a limited number of social attributes, pension contributions for individuals and the provision of unpaid care are reported, as well as savings held by households.

9.2.4 Panel Surveys

The final sources of attribute data to consider are two panel surveys led by the University of Essex; the British Household Panel Survey (BHPS) and Understanding Society. The BHPS ran continuously for 18 years from 1991 and ended in 2009. It had a sample size of 10,000 individuals aged 16 and over drawn from 5,000 households and comprised a household survey (administered to the household reference person) and individual questionnaires (Taylor 2010). The household questionnaire asked “questions about the accommodation and tenure and some household level measures of consumption” while the individual schedule covered a wide number of topics including neighbourhood, residential mobility, health and caring, current employment and earnings, employment changes over the past year, values and opinions and household finances and organization (Taylor 2010, p. 26-27).

Understanding Society is a longitudinal study of approximately 40,000 households in the UK which began in January 2009 with the overall purpose of providing “high quality longitudinal data about subjects such as health, work, education, income, family, and social life” (McFall 2010, p.1). Understanding Society replaces and incorporates the BHPS. The Understanding Society survey asks questions to all members of the sample household and initial results from wave 1 have recently been released. Alongside the census and Integrated Household Survey, Understanding Society has been identified by the UK Strategy for data resources for social and economic research 2009–2012, an initiative taken forward by stakeholders from ONS, the devolved administrations of the UK, government departments and private consultancies, as a key source of attribute information in the future (ESRC 2009).
10. CONCLUSIONS: AN INTEGRATED CONCEPTUAL FRAMEWORK

10.1 Summary of the framework and contents

The report has been organized in two parts. The conceptual framework for population and migration statistics in the UK represented the first part, where we focused on the context in which population migration statistics are set, the overall conceptual framework itself and a discussion of the issues concerning error and uncertainty. The second part of the report focused on data, processing and outputs, where we provided basic information about the main data types, followed by some details about how population, internal migration and international migration estimates are produced, and the sources of data underlying them. We also included a chapter on the important population and migration attributes, beyond age, sex and location of residence, that are needed for understanding and explaining societal change and the key sources for obtaining that information.

10.2 New developments and future plans

ONS is currently undergoing a transition in how it uses data. More access has been provided to administrative data available from other government departments, allowing them to take advantage of previously unavailable data sources. At the same time, pressure is being placed on ONS to reduce the costs and increase the efficiency of providing key statistics on populations. Traditionally, the census has provided the main benchmark for population estimation. In the future, there is the possibility that censuses will not take place, as they are expensive and time consuming, and some sort of quasi-population register will provide the main platform for estimating populations. The ONS Beyond 2011 Programme will be taking forward the work to develop, test and evaluate a range of options to deliver population benchmarks, as well as the socio-demographic information that the census also importantly delivers. ONS (2012d) provides links to a range of information about the Beyond 2011 Programme.

This work will be challenging because government departments must communicate, overcome confidentiality restrictions and work together in order to share data. The legal frameworks for doing so have only recently been tested. Furthermore, the administrative data sources currently available are not designed to track populations and their behaviours. They must be adapted and combined in order to fit the population concepts needed. To facilitate this communication and processing, a conceptual framework is needed to provide the foundation.

10.3 Concluding remarks

The production of population and migration statistics is complicated because populations are dynamic, heterogeneous, and influenced by the social and cultural environments, economic environments and natural and built environments in which the populations reside. In this report, we have presented a unified framework for population and migration estimates in the UK. The key elements are concepts, data, processing and outputs. Concepts are required to understand what types of population statistics are of interest and how the demographic components of change should be related to them. Concepts are also needed to match the measurements in the available data, which may come from censuses, surveys or administrative sources, to the statistic of need. As the available data are unlikely to match the concept exactly, and are very likely to contain error or miss out on certain groups of the population, processing is required.

The conceptual framework for UK population statistics presented in this document is useful for facilitating communication with users of population and migration statistics, and for ensuring that everyone understands the underlying concepts, the nature of the data available and the methods used to derive estimates of key statistics. One important message that we have focused on is the tension between concepts and outputs of migration and population statistics. There are many requirements and types of population statistics but no single source of information in the UK that covers all these requirements. The published outputs, therefore, are only able to address some of
the needs and are often produced by bringing together data from multiple sources, all with their relative strengths and weaknesses. Future work should consider extending the framework to include the relationships between different user needs and statistical releases.

In conclusion, we hope this report provides a foundation for understanding the relationships between the population concepts required by users and the data available to measure these in the UK, with some insights and clarifications on how migration and population statistics are produced in the UK. It should also provide a starting point for considering alternative data or methods for producing population and migration statistics. The report is intended to be adapted and improved as the production of population and migration statistics evolve.
# Glossary

<table>
<thead>
<tr>
<th>Terms</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>APS</td>
<td>Annual Population Survey</td>
</tr>
<tr>
<td>ASC</td>
<td>Annual Schools Census</td>
</tr>
<tr>
<td>CA</td>
<td>Council Area (district) in Scotland</td>
</tr>
<tr>
<td>CIDER</td>
<td>Centre for Interaction Data Estimation and Research</td>
</tr>
<tr>
<td>CMA</td>
<td>Census Metropolitan Area</td>
</tr>
<tr>
<td>DCLG</td>
<td>Department for Communities and Local Government</td>
</tr>
<tr>
<td>District</td>
<td>Local Authority, Unitary Authority, Council Area, Local Government District</td>
</tr>
<tr>
<td>CHI</td>
<td>Community Health Index</td>
</tr>
<tr>
<td>DWP</td>
<td>Department for Work and Pensions</td>
</tr>
<tr>
<td>EFS</td>
<td>Expenditure and Food Survey</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>Eurostat</td>
<td>The Statistical Office of the European Communities</td>
</tr>
<tr>
<td>Flag 4</td>
<td>Marker entered on new GP patient registration when previous address was outside UK</td>
</tr>
<tr>
<td>GHS</td>
<td>General Household Survey</td>
</tr>
<tr>
<td>GOR</td>
<td>Government Office Region</td>
</tr>
<tr>
<td>GP</td>
<td>General Practitioner</td>
</tr>
<tr>
<td>HA</td>
<td>Health Authority</td>
</tr>
<tr>
<td>HB</td>
<td>Health Board</td>
</tr>
<tr>
<td>HESA</td>
<td>Higher Education Statistics Agency</td>
</tr>
<tr>
<td>IHS</td>
<td>Integrated Household Survey</td>
</tr>
<tr>
<td>IMPS</td>
<td>Improving Migration and Population Statistics</td>
</tr>
<tr>
<td>IPS</td>
<td>International Passenger Survey</td>
</tr>
<tr>
<td>L2</td>
<td>Lifetime Labour Market Database</td>
</tr>
<tr>
<td>LA</td>
<td>Local Authority</td>
</tr>
<tr>
<td>LFS</td>
<td>Labour Force Survey</td>
</tr>
<tr>
<td>LGD</td>
<td>Local Government District in Northern Ireland</td>
</tr>
<tr>
<td>LLTI</td>
<td>Limiting Long-Term Illness</td>
</tr>
<tr>
<td>LTIM</td>
<td>Long-Term International Migration</td>
</tr>
<tr>
<td>MWS</td>
<td>Migrant Worker Scan</td>
</tr>
<tr>
<td>NHSCR</td>
<td>National Health Service Central Register</td>
</tr>
<tr>
<td>NI No</td>
<td>National Insurance Number</td>
</tr>
<tr>
<td>NIRA</td>
<td>Northern Ireland Statistics and Research Agency</td>
</tr>
<tr>
<td>NMD</td>
<td>New Migrant Databank</td>
</tr>
<tr>
<td>NPD</td>
<td>National Pupil Database</td>
</tr>
<tr>
<td>NRS</td>
<td>National Records Scotland</td>
</tr>
<tr>
<td>NS-SEC</td>
<td>National Statistics Socio-Economic Classification</td>
</tr>
<tr>
<td>OA</td>
<td>Output Area</td>
</tr>
<tr>
<td>ONS</td>
<td>Office for National Statistics</td>
</tr>
<tr>
<td>PAF</td>
<td>Postcode Address File</td>
</tr>
<tr>
<td>PBS</td>
<td>Points Based System</td>
</tr>
<tr>
<td>PC</td>
<td>Parliamentary Constituency in Northern Ireland</td>
</tr>
<tr>
<td>PCT</td>
<td>Primary Care Trust</td>
</tr>
<tr>
<td>PRDS</td>
<td>Patient Register Data System</td>
</tr>
<tr>
<td>S3RI</td>
<td>Southampton Statistical Sciences Research Institute</td>
</tr>
<tr>
<td>SCHI</td>
<td>Scottish Community Health Index</td>
</tr>
<tr>
<td>SMS</td>
<td>Special Migration Statistics</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>UKBA</td>
<td>UK Borders Agency</td>
</tr>
<tr>
<td>WRS</td>
<td>Worker Registration Scheme</td>
</tr>
</tbody>
</table>
REFERENCES


A conceptual framework for UK population and migration statistics


GeoPlace (2011) GeoPlace Spatial Address and Location Data. Online at http://www.geoplace.co.uk/.


A conceptual framework for UK population and migration statistics


A conceptual framework for UK population and migration statistics


APPENDIX: POPULATION ESTIMATES IN A POPULATION ACCOUNTING FRAMEWORK

The methods used to estimate the mid-year populations of national, local and small area populations derive from population accounts. Population accounts are matrices that show how populations move between starting and ending states. To these matrices are added vectors of totals which hold population and component totals, as long as the system of population and areas is comprehensive, as Stone (1971) demonstrated 41 years ago. The population accounts have different structures depending on the migration concept employed.

Table A.1 sets out the structure of population accounts using the movement concept. The table uses three spatial units, constituting the minimum system of interest needed. There is an area of interest, the rest of the country and the rest of the world. The “rest of” areas can be broken down into a set of sub-national areas and a set of countries or country groupings as required, so the specification is quite general. The rows of the table show the origin state of people in the time interval: either their residence immediately prior to the move or their location at first appearance in the system, i.e. at birth, which is recorded as mother’s place of usual residence at the time of maternity in the register of births. The columns of the table show the destination state of people in the time interval: either residence immediately after the move or location at final appearance in the system, that is, usual residence at time of death.

**Table A.1: The structure of a movement population account for all ages**

<table>
<thead>
<tr>
<th>Origin before move</th>
<th>Destination after move</th>
<th>Total deaths</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area of interest</td>
<td>Rest of country</td>
<td>Rest of the World</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>r</td>
</tr>
<tr>
<td>Area of interest</td>
<td>R1</td>
<td>M21</td>
<td>E2</td>
</tr>
<tr>
<td>Rest of country</td>
<td>M11</td>
<td>R2</td>
<td>E2</td>
</tr>
<tr>
<td>Rest of the world</td>
<td>I1</td>
<td>I2</td>
<td>0</td>
</tr>
<tr>
<td>Births</td>
<td>B1</td>
<td>B2</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>FP1</td>
<td>FP2</td>
<td>E1</td>
</tr>
</tbody>
</table>

Notes:
- \( R \) = residual or balancing term for area
- \( M \) = migrations (moves) from an origin area to a destination area
- \( E \) = emigrations (moves) from an origin to the rest of the world
- \( I \) = immigrations (moves) from the rest of the world to a destination area
- \( D \) = deaths in an area (usual residence of deceased)
- \( B \) = births in an area (usual residence of mother)
- \( SP \) = start population in a time interval from mid-year \( t \) to mid-year \( t+1 \)
- \( FP \) = end population in a time interval from mid-year \( t \) to mid-year \( t+1 \)
- \( 0 \) = flow not recognised, not involved in the country of interest
- \( T \) = total of flows into and out of country of interest
- \( + \) = summation over superscript replaced
- \( 1, 2 \) = superscripts for areas

The cells of the table identify population flows of interest. The letter \( M \) is used to represent migrations (moves) with the superscripts indicating location before and after the migration. The letter \( I \) is used to represent immigrations to the system of interest; the letter \( E \) indicates emigration from the system of interest. The letter \( B \) refers to births and \( D \) to deaths, as is conventional. The letter \( R \) appears in the diagonal of the matrix, referring to the residual population term necessary to
balance the account. As we shall see later, it drops out of the population estimation equation. This diagonal term is neither the number of survivors staying in the same area throughout the time interval nor the number of migrations internal to the area. The first of these interpretations is appropriate in transition population accounts while the second is appropriate when a model of migration between areas is being constructed. In the bottom right corner of the accounts matrix are zeroes, indicating that the system has been left open. We are not interested in vital events and moves that take place solely in the rest of the world. These cells make no contribution to the estimation of populations in the area of interest or in the rest of the country.

In the rightmost column of Table A.1 are shown the totals for inputs into the system of interest in the time interval. The term $SP$ refers to start population in the interval with a superscript indicating the area involved. The term $I^+$ is used to represent the total of immigration to the area of interest and the rest of the country. The term $B^+$ represents the total of births in the system of interest. In the bottom row are shown the totals for outputs from the system of interest in the time interval. The term $FP$ refers to final population in the interval with a superscript indicating the area. The term $E^+$ refers to the total of emigration from the system of interest and the term $D^+$ to the total of exits from the system through death. When we sum these total inputs we obtain a figure for total inputs, $T^+$. This should equal the sum for the total outputs from the system. If this is not the case then, there is a closure error for the system which must be resolved.

The interior table numbers add up to the marginal totals. This property follows from the following accounting identities. We use the letter $i$ to indicate a typical area of interest and the plus superscript indicates summation over the superscript replaced.

The variables in a row add to the start population:

$$SP^i = R^i + M^i + E^i + D^i.$$  \hfill (A.1)

The variables in a column add to the final population:

$$FP^i = R^i + M^i + I^i + B^i.$$  \hfill (A.2)

We compute the residual term, $R^i$, through subtraction:

$$R^i = SP^i - M^i + E^i - D^i.$$  \hfill (A.3)

Then we substitute the right-hand side expression for the term $R^i$ in accounting identity (A.2) and obtain:

$$FP^i = SP^i - M^i + E^i - D^i + M^i + I^i + B^i.$$  \hfill (A.4)

This is the accounting equation which underpins all rolled forward (in time) population estimates. At the end of each time interval we transfer the final populations of the time interval into the start of population of the next, using the time postscript as a period-cohort index:

$$SP^i(t+1) = FP^i(t)$$  \hfill (A.5)

and then supply fresh estimates of the change component terms for the next period. In practice, the populations are disaggregated by age and sex because the change components vary systematically across both sexes and ages. In the next section we describe the rolled forward equations for populations by age and sex.

The first time interval to which the population estimation equations apply starts with the census and covers the period between the latest decennial census and mid-year. In 2011 this interval will be between 27 March 2011 and 30 June 2011. Then the estimates rolled forward one year at a time from 30 June in the start year and 30 June in the next year.
To estimate populations by age and sex, it is necessary to understand the age-time plan used in the method and the age-time plans used in other models and for which data are prepared. Age-time plans are normally termed Lexis diagrams after their German 19th Century promoter. However, since Vanderschrick (1992) casts doubt on Lexis's provenance as the inventor and because the version now used was developed by the French Demographers Louis Henry and Roland Pressat (Pressat 1969), the more general term "age-time plan" is used here.

In Figure A.1, four age-time plans are presented, along with their links to applications. The four graphs have common axes: age is represented on the vertical axis and time on the horizontal axis. The shaded areas represent different ways in which demographic event data can be captured. The period-age scheme records vital events in a time period \( (t \to t+1) \) and within an age \( (x \to x+1) \). This is used by ONS to report vital events including migrations (moves). A more informative age-time plan is the age-period-cohort in which vital events are classified by time period, age at which the event takes place and the birth cohort experiencing the event. This is in standard use in many European countries but only occasional use up to the present by ONS. If the individual event record holds both the date of the event and the date of birth of the individual experiencing the event, then it is an easy matter to tabulate vital events using the age-period-cohort scheme. The advantage of this method of reporting demographic events is that they can be reassembled (precisely) for input to the two main demographic models, the cohort-component model which uses the period-cohort age-time plan and the life table model which uses the age-cohort time plan (see Figure A.1). If only period-age tables are available, unsatisfactory approximations must be made in preparing input data for both cohort-component models and life table models.

Figure A.1: Four age-time plans used with population statistics
When we develop population estimation equations for ages, we use the period-cohort scheme but need to distinguish three different cases, which are shown in Figure A.2: the typical period-cohort (the left hand graph), the first or new-born period-cohort (the middle graph) and finally the last period-cohort (the right hand graph).

In the typical period-cohort, people survive from being age \( x \) at time \( t \) to being age \( x+1 \) at time \( t+1 \). In the new-born period-cohort, babies cross the age 0 line in the time interval \( t \) to \( t+1 \) and then survive to being in age 0 at the end of the period. The final period-cohort, represented by the area shaded grey in the rightmost graph, is open-ended. If birthday \( z \) is the threshold of the last age, people in the last period-age move from being aged \( z+1 \) at time \( t \) to being aged \( z+1 \) at time \( t+1 \). To compute the starting population for period \( t+1 \) to \( t+2 \) we need to add together the final populations of this period-cohort (shaded purple in the graph) with that of the last but one. The threshold used for the last age has moved upwards over past decades in ONS statistics reflecting steady life expectancy improvement. It is likely to move upwards to age 100 in most age related series over the next decade.

![Figure A.2: Age-time diagrams showing how the period-cohort estimates work](image)

The population accounts associated with the typical period-cohort are set out with the relevant algebraic variables in Table A.2 and for the new-born period-cohort in Table A.3. The accounts tables have one fewer row that the all ages table (Table A.1) because births are handled in the new-born period-cohort table. The algebraic notation used in the tables adds a subscript \( x \), which refers to the period-cohort that starts at age \( z \) at time \( t \). This subscript applies to all variables and only changes when final populations from one period-cohort are used as start populations in the next period-cohort.
Table A.2: The structure of movement population accounts for a typical period-cohort

<table>
<thead>
<tr>
<th>Origin before move</th>
<th>Area of interest</th>
<th>Destination after move</th>
<th>Deaths</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>r</td>
<td>d</td>
</tr>
<tr>
<td>Area of interest</td>
<td>$R_x^1$</td>
<td>$M_x^{21}$</td>
<td>$E_x^1$</td>
<td>$D_x^1$</td>
</tr>
<tr>
<td>Rest of country</td>
<td>$M_x^{21}$</td>
<td>$R_x^2$</td>
<td>$E_x^2$</td>
<td>$D_x^2$</td>
</tr>
<tr>
<td>Rest of the world</td>
<td>$I_x^1$</td>
<td>$I_x^2$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>$FP_{x-1}^1$</td>
<td>$FP_{x-1}^2$</td>
<td>$E_x^+$</td>
<td>$D_x^+$</td>
</tr>
</tbody>
</table>

Notes: (1) $x$ = start age for persons in a time interval who end the time interval in age $x+1$; (2) period-cohort = age-time space for one time interval and one birth cohort.

Table A.3: The structure of movement population accounts for the new-born period-cohort

<table>
<thead>
<tr>
<th>Origin before move</th>
<th>Area of interest</th>
<th>Destination after move</th>
<th>Deaths</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>r</td>
<td>d</td>
</tr>
<tr>
<td>Area of interest</td>
<td>$R_{-1}$</td>
<td>$M_{-1}^{12}$</td>
<td>$E_{-1}$</td>
<td>$D_{-1}$</td>
</tr>
<tr>
<td>Rest of country</td>
<td>$M_{-1}^{21}$</td>
<td>$R_{-1}$</td>
<td>$E_{-1}$</td>
<td>$D_{-1}$</td>
</tr>
<tr>
<td>Rest of the world</td>
<td>$I_{-1}$</td>
<td>$I_{-1}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>$FP_{-1}^1$</td>
<td>$FP_{-1}^2$</td>
<td>$E_{-1}$</td>
<td>$D_{-1}$</td>
</tr>
</tbody>
</table>

Notes: -1 = start age for persons born in a time interval who end the time interval in age 0

The cohort-component version of the population estimation equation is as follows:

$$FP_x^i = SP_x^i - M_x^{it} - E_x^i - D_x^i + M_x^{it} + I_x^i.$$  (A.6)

Before the estimation equation for the next period is implemented, the final population of period-cohort $x$ in period $t,t+1$ is transferred into the start population of cohort $x+1$ in period $t+1,t+2$:

$$SP_{x+1}^i = FP_x^i,$$  (A.7)

except for the final period-cohort where two final period-cohort populations are merged:

$$FP_{x+1}^i = FP_{x+1}^i + FP_{x+1}^{i-1}.$$  (A.8)

The new-born period-cohort, labelled here as “-1”, uses births as its start population:

$$FP_{-1}^i = B_{-1}^i - M_{-1}^{it} - E_{-1}^i - D_{-1}^i + M_{-1}^{it} + I_{-1}^i.$$  (A.9)

If we recognize that internal or international migrations come from lots of different destinations, then the population estimation equation expands to include origin-destination movements in a multiregional or multistate demography framework (Rogers 1995; Schoen 2006), as follows:
\[ FP_x^i = SP_x^i - \sum_{j \neq i} M_{x}^{ij} - \sum_{k} E_{x}^{k} - D_{x}^{i} + \sum_{j \neq i} M_{x}^{ji} + \sum_{k} I_{x}^{k}, \]  \hspace{1cm} (A.10)