Frameworks for Guiding the Development and Improvement of Population Statistics in the United Kingdom

James Raymer[1]*, Phil Rees*[2]*, and Ann Blake*[3]

The article presents central frameworks for guiding the development and improvement of population statistics. A shared understanding between producers and users of statistics is needed with regard to the concepts, data, processes, and outputs produced. In the United Kingdom, population estimates are produced by conducting decennial censuses and by estimating intercensus populations through the addition and subtraction of the demographic components of change derived from registers of vital events and from a combination of administrative data and surveys for internal and international migration. In addition, data cleaning, imputation, and modelling may be required to produce the desired population statistics. The frameworks presented in this paper are useful for aligning the required concepts of population statistics with the various sources of available data. Taken together, they provide a general 'recipe' for the continued improvement and expansion of official statistics on population and demographic change.

Key words: Population statistics; frameworks; data sets.

1. Introduction

Population statistics are used by national governments to distribute money across local governments for managing services, such as healthcare and education. They are used to establish the boundaries for political constituencies and to provide denominators for other measures, such as fertility or unemployment rates. They are used to manage and plan for future water, power, and sanitation needs. They can also be used to decide the number and types of homes to be built and design centres for shopping and leisure activities. Companies use population statistics to target their goods and services at specific groups of people (Boyle and Dorling 2004) and for workforce planning purposes. In other words,

[1] Australian National University, School of Demography, ANU College of Arts and Social Sciences, 9 Fellows Road, Acton Act 2601, Australia. Email: james.raymer@anu.edu.au
[2] University of Leeds, School of Geography, University Road, Leeds, West Yorkshire LS2 9JZ, UK. Email: p.h.rees@leeds.ac.uk
[3] Office for National Statistics, Segensworth Road, Fareham, Hampshire PO15 5RR, UK. Email: ann.blake@ons.gsi.gov.uk

population statistics are central for understanding society and societal change and are widely used in comparisons with other countries and areas within countries (Stone 1971; United Nations 1975).

Published outputs of population statistics are the result of matching available data to particular needs or concepts. The concepts are driven by user demand. In some cases, the data are processed or combined with other information to make statistics. Furthermore, the published statistics may not meet all of the needs of various user groups. As this article highlights, there are many requirements and types of population statistics. However, there are rarely single sources of information that cover a wide variety of needs. Instead, population statistics bring together data from many sources, all with their relative strengths and weaknesses.

The production of population statistics is further complicated because populations are both dynamic and heterogeneous. They change continuously according to the addition of births, subtraction of deaths and addition or subtraction of migrants. 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 (Bycroft 2011). In order to understand population statistics, one must first realise that they only represent a 'snapshot' of a population at a particular time or a flow between two fixed time points.

With such a demand and need for information on populations and their demographic 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. It is therefore vital not only that the statistics are reliable and robust, 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 information needs to be publicly available, transparent and understandable.

To help with the task of assembling population statistics for a country, national statistics offices require frameworks or agreed sets of concepts and methods, aligned to those available through current collection instruments and databases. Shared understanding of the frameworks within an organisation and across users facilitates the production and improvement of population statistics. In this article, we show how frameworks for population statistics can provide the main basis for achieving this. We use the United Kingdom (UK) as the illustration, but the frameworks may be generalised or adapted to other countries and data systems.

2.   Frameworks for Population Statistics

In 1971, Richard Stone defined a system of social and demographic statistics, which he then refined for the United Nations in 1975. The system covered the whole range of government statistics, starting 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 introduced the

concept of the life course, which is so important in contemporary social science research. Stone's system represents an important and basic foundation for how we think about official statistics today. Our frameworks presented in this article focus on demographic stocks and flows, which form the basic elements of the wider design set out by Stone.

In this section, we first present a framework for official statistics and then illustrate how this can be applied to produce statistics on the usual resident population. We also present a framework for the underlying mechanism of population change. The frameworks are designed to be general across the many different ways of producing population statistics. For illustration, we describe how they are being used by the United Kingdom's Office for National Statistics (ONS) in their Beyond 2011 Programme, which is "taking a fresh look at options for the production of population and small area socio-demographic statistics for England and Wales" (ONS 2013a). This includes exploring options based on administrative data (ONS 2013b).

2.1. Official Statistics

In designing a framework for population statistics, it is useful to first think about the activities and stages involved in producing official statistics (Laux 2002). In Figure 1, the functions of Society, Concepts, Data, Processing, Outputs, and Validation are presented. *Society* represents the economic and social conditions of the country producing the statistics. It determines the type of information required. *Concepts* refer to particular statistics of interest, such as the usual resident population, rates of unemployment, welfare provision or persons present without citizenship. *Data* are any information gathered concerning the required statistic, usually obtained from censuses, surveys or administrative registers. *Processing* represents data cleaning, imputation, combining two or more information sources through matching or proportioning, and statistical modelling to ensure that the data more closely match the required concept. *Outputs* are the published statistics or estimates. *Validation* is the procedure of assessing the quality of the published statistics in relation to the concepts required, often resulting in periodic revisions or improvements in data collection or processing.
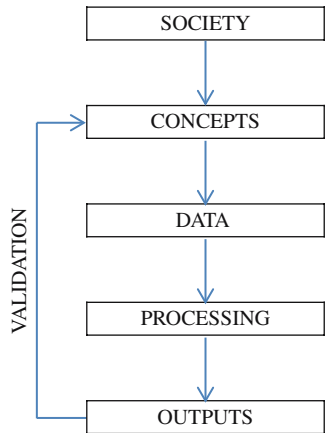

Fig. 1. A framework for producing official statistics

2.2.   *Population Statistics*

The general framework for official statistics presented in Figure 1 may be applied to the production of population statistics. Many types of population statistics can be produced, such as the population present at the time of a census or survey (*de facto* population), the population considered permanent or usually resident (*de jure* population), the population considered temporary (tourists, business travellers, short-term migrants), the population born in the country (native born) and the population born abroad (foreign born). We focus on the statistics that form the basic requirement for most national statistical offices, namely the usual resident population by age and sex, living in area i at time t. Adopting the United Nations (2007) definition, usual residents are all persons who reside, or intend to reside, in a place continuously for either most of the last twelve months or for twelve months or more. This includes nationals, foreigners, undocumented persons, applicants for asylum and refugees. Counts of usual residents provide ". . . the best indication of where people will demand and consume services, and . . . is therefore most relevant for planning and policy purposes" (United Nations 2007, 132).

To illustrate how the framework for official statistics can be applied in the production of statistics of usual residents by age, sex and local authority in the United Kingdom, consider the expanded framework presented in Figure 2. Here, more details are provided under the main headings of society, concept, data, processing, and output. Validation is not included in the diagram, but of course it is necessary for the continued improvement of the population statistics over time. We also include two columns covering background and other considerations, which may be interpreted as the context and specific factors for a particular society, respectively.

There are many issues to consider in the production of population statistics, which are largely driven by the need for distributing national resources, planning, and social welfare. What makes the UK different from other countries is its economy and culture, and this varies over time depending on, for example, the political climate and available resources. The UK does not have a population register or common identification number or code system (Poulain et al. 2006, 112–113), and has historically relied on decennial censuses to produce population statistics, with demographic estimation (described in Subsection 2.3) used to produce statistics for years between censuses. Note that care must be taken, when comparing censuses, to allow for coverage differences as well as conceptual differences in defining the population of interest. Very often, to aid these comparisons, several different population bases will be provided by national statistical offices for the basic population counts.

As mentioned previously, at any time, there are always several populations that may be measured or conceptualised. In the UK, a midyear (30 June/1 July) 'usual resident' population estimate is produced. This statistic is recommended by the United Nations for international comparability, although the practicalities of identifying usual residents may differ from country to country. Furthermore, the size and characteristics of the population may vary greatly, depending on both the time of day and day of the year measured. The usual resident population represents a 'night-time' measure, which captures the population where it sleeps. A 'day-time' measure captures the population where it goes to school, to work, to shop or to pursue leisure activities. Finally, it is important to be clear about the
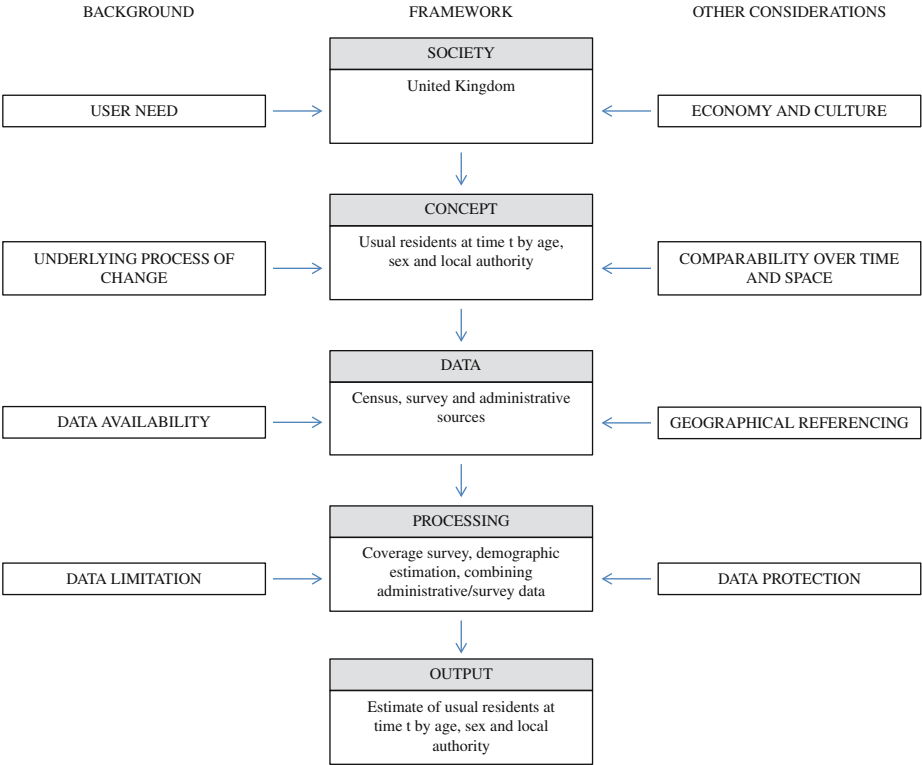
BACKGROUND   FRAMEWORK   OTHER CONSIDERATIONS

SOCIETY

United Kingdom

| USER NEED | → | | ← | ECONOMY AND CULTURE |

CONCEPT

Usual residents at time t by age, sex and local authority

| UNDERLYING PROCESS OF CHANGE | → | | ← | COMPARABILITY OVER TIME AND SPACE |

DATA

Census, survey and administrative sources

| DATA AVAILABILITY | → | | ← | GEOGRAPHICAL REFERENCING |

PROCESSING

Coverage survey, demographic estimation, combining administrative/survey data

| DATA LIMITATION | → | | ← | DATA PROTECTION |

OUTPUT

Estimate of usual residents at time t by age, sex and local authority

Fig. 2. A framework for producing population statistics in the United Kingdom

criteria for the inclusion or exclusion of individuals. Business travellers and visitors are usually excluded from official and international statistics on population and migration. Temporary workers may be included in official estimates of short-term migrants, but not in the usual resident population.

Regardless of the population concept being measured, it can only ever represent a snapshot of the continuous process of population change. This is further illustrated in Figure 3, which shows how populations present in an area may change over time. If the interest is tracking the changes of usual residents, each of the entry and exit components needs to be defined accordingly. Here, for example, only those who change their country of usual residence would be included as international entries or exits. All other international entries and exits would be excluded from the official statistics. Likewise, the

Actual population at time *t-n*
Usual residents
Migrants
Temporary workers
Business travellers
Visitors

Entries between *t-n* and *t*
Births
International moves
Internal moves

Exits between *t-n* and *t*
Deaths
International moves
Internal moves

Actual population at time *t*
Usual residents
Migrants
Temporary workers
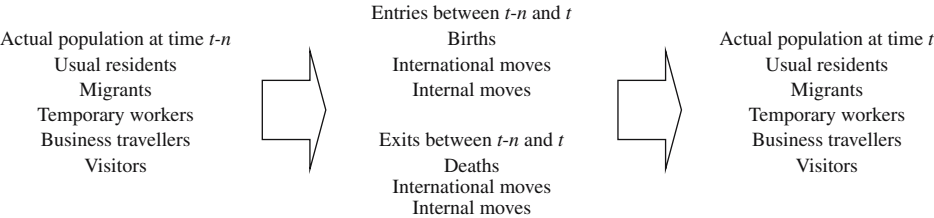Business travellers
Visitors

Fig. 3. The dynamics of actual population change for a geographic location within a country

births, deaths and domestic movements of temporary workers, business travellers and visitors would also be excluded.

There are two ways in which estimates of the usual resident population can be obtained: (1) using a source or combination of sources that can be used to count the population at a particular point in time and (2) combining sources of information on population stocks (e.g., recent census) with information on demographic events over time (see Subsection 2.3). The sources of demographic data may be censuses or surveys, or may arise from administrative processes including the registration of births and deaths or the need to access healthcare or state benefits. Each data source has limitations with regard to collection and storage. This usually concerns the amount of detail included and whether it contains individual information or an aggregation of individual information. It also concerns the amount of characteristic detail included in the data. Typically, censuses and surveys include much characteristic information for understanding societal differences and change, whereas administrative data contain only basic information necessary for operational purposes (see Subsection 2.4).

It is difficult to design data sources to measure particular population concepts and to capture all of the population required. Moreover, in the case of administrative data, population measurement is not the primary purpose. Thus, national statistical offices often use additional methods to improve the data, aligning the information with the required concept. In the UK, for example, coverage surveys and imputation are used to produce the census estimates (ONS 2013c). For estimates between census years, demographic event data are combined with statistical estimation used to augment the data on internal and international migration. Finally, there is an increasing need to provide measures of uncertainty – or, conversely expressed, of accuracy – with the estimates so that they may be interpreted correctly (see Section 3). The outputs or published statistics are estimates, since they are only able to come close to the 'true' conceptual measure but do not match it exactly. These statistics are also published with supporting information about how they have been produced, allowing the user community to make informed judgements on the current state of the population and its likely future.

To summarise, it is important to understand the various factors involved in producing population statistics. The framework for population statistics starts with the societal context and extends to the required published outputs. Next, we describe how the usual resident population can be aligned with the underlying mechanism for population change, involving the demographic accounting equation.

2.3. *Underlying Mechanism of Demographic Change*

There are two ways that population statistics can be produced: enumeration and demographic accounting. In practice, both are often used to understand the population change occurring and to verify the quality of the estimates being produced. In the UK, censuses are used to provide accurate estimates of the population every ten years. In between censuses, midyear estimates are produced by rolling forward the age- and sex-specific census estimates based on the number of births, deaths, domestic migrations and international migrations that occur within each year utilising the demographic accounting equation (see Figure 4 and equations below). To maintain an accurate picture of population
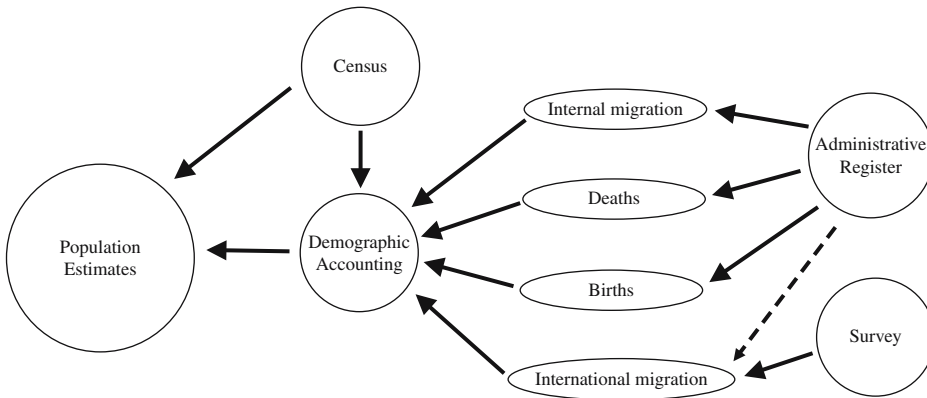
Fig. 4. *Overview of population estimation in the UK*

change over time, care must be taken to match the model used for demographic estimation with the nature of the data available. Note that for the production of international migration statistics in the UK, both survey (major source) and administrative (minor source) data are used (represented by solid line and dashed line arrows, respectively).

When producing estimates based on demographic accounting, it is important to take into consideration the close dependence between the migration definition and the demographic accounting model. There are two types of migration data: events and transitions. Events refer to the number of moves that occur within a particular time period, whereas transitions refer to data collected on places of residence at two points in time. The demographic accounting model essentially rolls forward the population estimate from the last reliable estimate (e.g., census) to successive years until the next reliable estimate, where the error of the process can be assessed. The same demographic accounting process is also used for estimating future population totals (i.e., population projections), except that observations on fertility, mortality and migration are replaced with estimated future (projected) figures.

Demographic change accounts based on events are presented in Table 1. These accounts represent an extension of those originally developed by Stone (see United Nations 1975). The variable M_{ij} represents internal migration events from one area i to another area j. The n subscript denotes the number of areas. Note that we ignore internal migration events when $i = j$. Instead, we enter terms R_i, which are accounting balances, the result of subtracting from the start population all possible exit events. Total (internal) outmigrations from each area are denoted by M_{i+} and total (internal) inmigrations to each area are denoted by M_{+i} (see the Table 1 notes for explicit definitions). Here we use the subscript i as a general index for a region. The I_i variable signifies the number of immigration events from outside the system of interest and E_i tabulates the corresponding emigration events. The vital events of death, D_i, and birth, B_i, complete the flows in the table. The sum of the numbers in the rows adds up to the populations at the beginning of the time interval, $P_i(t)$. The balancing term is obtained by subtracting the total number of outmigrations, emigrations and deaths from the population at the beginning of the time interval, that is,

$$R_i = P_i(t) - M_{i+} - E_i - D_i. \tag{1}$$

Table 1. *A demographic accounting framework for population statistics*

From origin:	To destination: Area 1	Area 2	. . .	. . .	Area i	. . .	. . .	Area n	Emigrations	Deaths	Total
Area 1	R_1	M_{12}	. . .	. . .	M_{1i}	. . .	. . .	M_{1n}	E_1	D_1	$P_1(t)$
Area 2	M_{21}	R_2	. . .	. . .	M_{2i}	. . .	. . .	M_{2n}	E_2	D_2	$P_2(t)$
.	.	.			.			.	.	.	.
Area i	M_{i1}	M_{i2}	. . .	. . .	R_i	. . .	. . .	M_{in}	E_i	D_i	$P_i(t)$
.	.	.			.			.	.	.	.
Area n	M_{n1}	M_{n2}	. . .	. . .	M_{ni}	. . .	. . .	R_n	E_n	D_n	$P_n(t)$
Immigrations	I_1	I_2	. . .	. . .	I_i	. . .	. . .	I_n	0	0	I_+
Births	B_1	B_2	. . .	. . .	B_i	. . .	. . .	B_n	0	0	B_+
Total	$P_1(t+1)$	$P_2(t+1)$	. . .	. . .	$P_i(t+1)$	. . .	. . .	$P_n(t+1)$	E_+	D_+	

Definitions of variables and subscripts: P = population, R = balancing terms, M = internal migrations (within a country), E = emigrations, I = immigrations, B = births, D = deaths, 0 = structural zeroes, t = time, i = subscript for area, n = number of areas and $+$ = summation over areas. Total internal outmigrations from area i = $\Sigma_{j=1 \text{ to } n, j \neq i}$ M_{ij} = Total internal inmigrations to area i = $\Sigma_{j=1 \text{ to } n, j \neq i}$ M_{ji} = M_{i+}. Total internal inmigrations to area i = $\Sigma_{j=1 \text{ to } n, j \neq i}$ M_{ji} = M_{+i}.

The variables in the columns of Table 1 add up to the populations at the end of the time interval. We can compute these by adding to the balancing term the total inmigrations, immigrations and births, that is,

$$P_i(t+1) = R_i + M_{+i} + I_i + B_i. \tag{2}$$

If we combine these two equations, the balancing term cancels out, and we obtain the familiar components of the population change equation:

$$P_i(t+1) = P_i(t) - M_{i+} - E_i - D_i + M_{+i} + I_i + B_i. \tag{3}$$

Although the demographic accounting model is applied similarly, it is important to understand how the movement and transition concepts of migration differ. To illustrate, consider Figure 5, where the vertical axis represents space and is divided into two regions and 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.

The demographic accounts in Table 1 are specified for the whole population (all ages). When the population is rolled forward from one year to the next, the accounts and associated population change equations can be specified for each age, using period cohorts. The only difference is that the birth terms are omitted from the accounts and instead are used as the starting population in the accounts for the new-born period cohort. Furthermore, Table 1 is specified for one country with several subnational regions and one
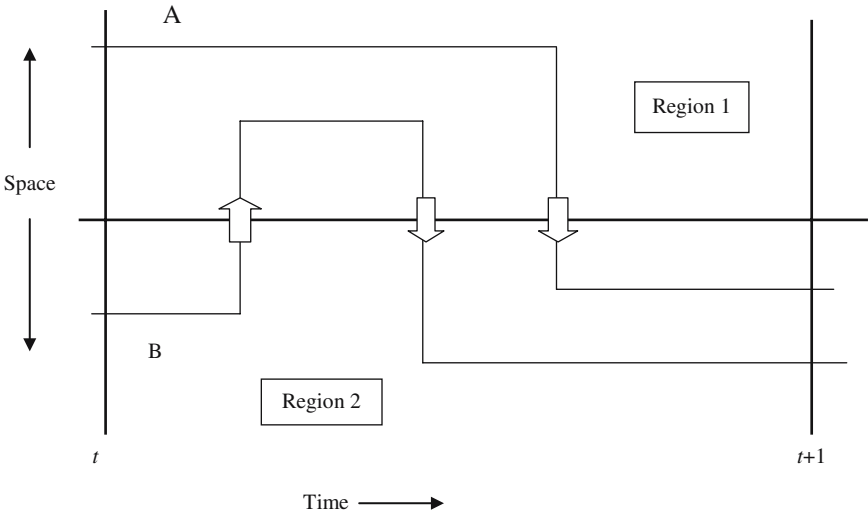

Fig. 5. A time-space diagram illustrating different migration measurement concepts

external region, the rest of the world. However, this spatial system can be expanded to include a large number of areas. The system can also be collapsed to estimate the national population change. Furthermore, the demographic accounting framework could also designate internal regions as separate countries (e.g., the 28 members of the European Union) or a combination of both national and subnational units can be employed (see, e.g., Kupiszewski and Kupiszewska 2008).

If an end of interval population, $P_i(t + 1)$, is independently available from a population register or an end-of-decade census, then the error of estimation may be calculated. To reconcile population change through the interval with the start and end population stocks, some corrections must be made to the component terms using assessments of the likely errors involved. To estimate the population at time $t + 1$, the demographic accounting model therefore requires a population stock at time t and information about births and deaths (natural increase measures) and domestic and international migration between time t and $t + 1$.

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 sex and birthplace of the child and the age and residential location of the mother. Deaths are recorded for all persons by age, sex and residential location. Migration data, on the other hand, are obtained from general-purpose censuses, surveys or administrative registers. Unlike fertility or mortality, the practical measures of migration obtained from these sources often do not coincide with theoretical or contextual definitions of migration (Bell et al. 2002). For example, Raymer and Smith (2010, 703) describe migration as "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." Migration can involve people moving within a country and within localities, as well as across international borders. The factors that separate migration from other forms of mobility (e.g., daily commuting, weekday/ weekend commuting, holiday visits or seasonal moves) are generally the distance travelled and the length of time spent in the destination (or away from the origin). In practical terms, migration can be defined as relocations between administrative areas and mobility as relocations within areas. Intra-area migration is not required for population change accounts (Table 1). 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 that may have relevance for the estimation of the population of very small areas, such as output areas or postal sectors.

2.4. Population Characteristics

Users of population statistics are interested in the characteristics of the population and how those characteristics change over time. Age and sex are considered the baseline characteristics required for population statistics because many other attributes have a close relationship with them. For understanding change or differences between population groups, it is useful to have more detailed attribute information, depending on the need or users. For example, for those interested in the integration of immigrants, 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 provision of services, information on population health, number of children and economic activity are useful. The availability of data on characteristics will depend on the type of information required and the data available to meet the need.

2.5. Geographic Classifications

Geographic classifications are fundamental for understanding society and population change. There are many different ways of representing geography, depending on the location information available in the data source. Statistics are usually produced for local-authority districts, counties and regions in England and Wales, and 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, workforces, as well as for comparing different spatial patterns of residence according to ethnicity or density.

3. Output Uncertainty

Information about uncertainty in the published statistics can help users to gain a better understanding and use of the statistics. If the sources of uncertainty are known, they 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. For example, 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 responded by developing methods for measuring error and defining confidence intervals for the England and Wales local authority midyear population estimates (ONS 2012b). They also developed a table of key indicators, which can be used to identify local authorities with characteristics associated with greater likelihoods of uncertainty in their midyear estimates, such as the proportion of students present.

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 errors. For example, in the UK, internal migration and international migration are considered the most problematic in terms of population error. Registrations of births and deaths, on the other hand, are considered highly reliable since they have a very clear legal framework and long history of data collection. The degree of uncertainty in the time series of population estimates is most often revealed when the results of the next census become available and often, as a result, the annual population estimates are revised to coincide with both the most recent and the previous census. To achieve a fully consistent time series of population change, it may also be necessary to revise the components of change. The vital statistics (births and deaths) normally do not need revision, though the associated demographic rates are often revised as a result of changes to the population at risk.

Within the area of survey methodology, the framework for survey error provided by Groves et al. (2009) provides a useful basis for considering the errors underlying data collected by surveys. The framework is also useful when considering data generated by a
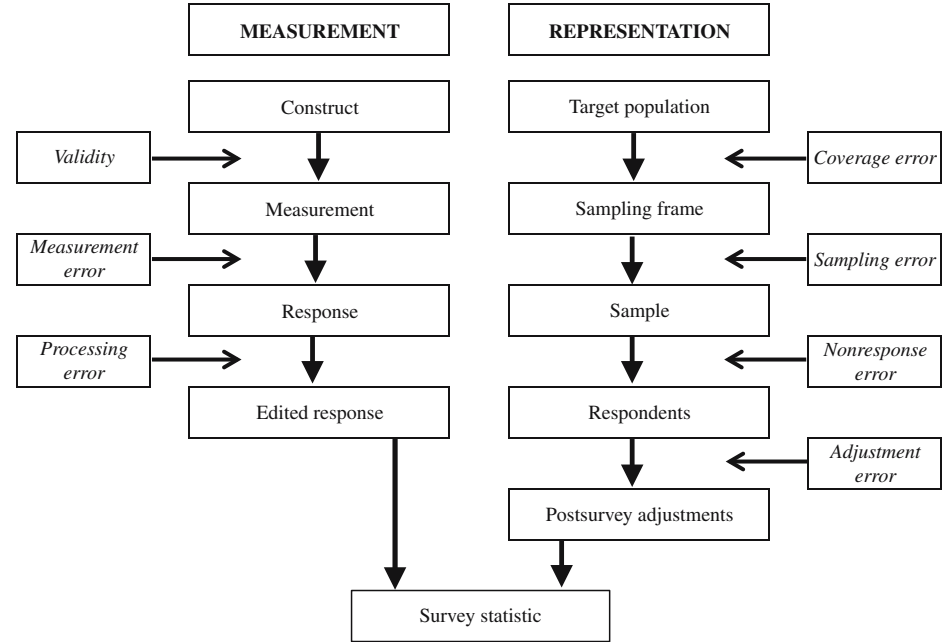
Fig. 6.   A framework for survey error. Source: Adapted from Figure 2.5 in Groves et al. 2009, 48

census or administrative system. Figure 6 sets out the framework, showing the processes used in gathering population data and where errors arise. Errors may be systematic or random, and they affect the uncertainty of the statistics in two different ways: systematically or randomly. An important purpose of including uncertainty measures in official statistics is to prevent users from making inferences that are not supported by the data.

3.1.   Accuracy

Accuracy refers to the closeness of the estimates to the true values. When producing population estimates, we know that accuracy is greater for larger populations as it takes more to influence their population change. We also know that accuracy decreases the further away in time the estimates are from a census. For example, in the UK, we would expect the 2012 population estimates to be considerably more accurate than the 2020 population estimates will be when they are produced, because the most recent (available) census occurred in 2011.

Bias, which is caused by systematic errors, refers to whether the estimates have higher or lower values on average in comparison to the true values. The sources of bias include both those that are known beforehand and those that appear unexpectedly. In general, the known biases are those that involve areas that are difficult to count, such as those with highly mobile populations, including students, migrants, homeless persons or armed-forces personnel, and those with high deprivation, unemployment or crime. These areas contain populations that are less likely to fill in questionnaires or register in an administrative source. 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 aged 15–29 years (Raymer et al. 2011, 80–81). In 2007, there were 55 per cent more females in the 20–24 year old age group changing their General Practitioner 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 or failures in deregistration. Finally, unexpected biases may result from a particular methodology used to estimate the population totals or to a change in the way data are collected. For example, when net migration rates are used in demographic accounting models instead of origin-destination-specific rates, there is a tendency to overestimate areas of growth and underestimate areas of decline (Rogers 1990).

Information on the potential sources of inaccuracy (due to systematic or random errors, or both) 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 with more or less uncertainty. These two aspects of uncertainty – size and causes – can also point to areas where methodology could be improved or where additional data should be gathered. They also provide a more realistic picture of population change. Furthermore, as national statistics offices continue to improve their methodology for estimating populations, the availability of measures of uncertainty could inform the extent of the improvement. The actual measures of uncertainty may vary from probabilistic predictive intervals to summary statistics, for example, Mean Algebraic Percentage Error for systematic deviation (bias) or Mean Absolute Percentage Error as a measure of precision, or spread, due to random variation. Finally, difficulties may arise in calculating the uncertainty measures, especially for population estimates where no information exists on the true values (e.g., postcensus population estimates and projections). Here, statistical models may be used to estimate the uncertainty.

3.2. 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 the lives of most people. Second, many data sources are not designed to capture migration specifically; rather, migration measures are a by-product obtained by additional analyses of administrative or general purpose 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 be overcome:

1. Difference between what the available data measure and what is required by the population estimation model (Rees 1985).
2. Small sample size of most survey data sources. 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.
3. Coverage differences between the populations targeted in survey and administrative sources and the coverage needed in the model for generating population estimates or projections.

4. 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.

5. 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.

What consequences does a migration measure have for population estimation? Rees (1985) points out that you need to match the concept used to measure migration with the concept used in the population estimation equation. The standard estimation equation assumes that migration is measured as a count of relocation events alongside the event counts of births and deaths. It is possible to develop a population estimate equation that uses migration measured as transitions, but the model is complicated for the UK for two reasons. First, there is no population register from which transitions can be counted for the population as a whole. In the UK, a census is administered only once in ten years and the main migration question captures transitions only for the year prior to the census. Second, the system of population statistics used by ONS is based on the movement concept using data on events. This is why ONS applies adjustment factors to the Patient Register Data System (PRDS)-derived migration measures which are based on the transition concept. The adjustment ratios are computed by comparing counts of moves (migrations as relocation events) between Former Health Authorities in a data set based on the NHS Central Register against the counts of transitions aggregated appropriately from the PRDS system (ONS 2012c).

3.3.   The Challenge of Measuring Special Populations: The Armed Forces and Prisoners

The population estimation methods described above depend on being able to measure all demographic components of change for all population groups. This is often true for births and deaths, where the registration has a legal basis. For the internal and international migration components, some populations may be left out. For instance, the NHS register does not report the movements of the Armed Forces or the prison population. The current approach to estimating populations is to subtract, at the start-of-time interval, the Armed Forces population and the prisoner population 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 and the Ministry of Justice, respectively.

4.   Applications of the Population-Statistics Framework

In this section, we illustrate how the frameworks presented in Section 2 can be applied to local area immigration statistics and to combine various administrative data sources to measure the usual resident population.

4.1. Immigration to Local Authorities in the United Kingdom

As an illustration of how the population statistics framework may be applied, consider the user requirement for the population within a particular local authority in 2010. The need could be conceptualised as "the usual resident population within local authority *j*, 2010". Here, data are required that match the concept of "all migrants arriving in the UK and residing for at least a year in local authority *j*, mid-2009 to mid-2010" to allow estimation of the population. Next, assume that there are only two sources of information available: migrants identified in the International Passenger Survey (IPS) and National Insurance Number (NINo) registrations to non-UK nationals. As shown in Figure 7, the IPS identified 536 (95% confidence interval: 506 to 566) thousand persons coming from abroad between mid-2009 to mid-2010 who intended to stay at least twelve months (ONS 2013g) (Note that this estimate of immigration was revised upwards after the 2011 Census to 579 thousand). 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 period, there were 668 thousand new NINo applications recorded for foreign citizens aged 16 years or over. 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 intending to stay for at least a year." The main advantage of the NINo registration is that it measures those "residing in local authority *j*." Neither of these sources alone can provide data that meets the concept above. Additional information about the number of visitor and migrant switchers (i.e., those changing their intentions) for the IPS data and the lag between arrival and registration and the duration of stay for the NINo data would be required. The two data sources can be aligned by separating the foreign citizens aged 16+ years from the rest of immigrants in the IPS. It is then possible to produce estimates of "foreign citizens aged 16+ years arriving in the UK for at least a year residing in local authority *j*, 2010." This illustration shows how the framework may be used to take a user requirement and align the available data to the concept needed; it also shows the obstacles and limitations that often occur with such a process.

Figure 7 represents an illustrative example based on the more complex estimation process described in ONS (2011) based on work by Boden and Rees (2010), which has improved estimates of immigration to local authorities in England and Wales. The method divides the national estimate of immigration into different flows by purpose of immigration – for work, for study, as returning citizens or residents, and as dependants. Different administrative sources are used to distribute each "purpose group." For example, higher-education records, which cover all newly registered foreign students, are used to estimate the local-authority distribution of foreign student arrivals. This method replaces a less accurate process that hierarchically distributed yearly IPS immigration totals: (1) into regions according to three-year averaged Labour Force Survey estimates; (2) into 'intermediate' geographies based on IPS estimates; and (3) into local authorities based on the previous Census (2001) counts.

4.2. Combining Administrative Data to Estimate Usual Residents by Age and Sex

In this subsection, we describe how ONS is exploring the quality and application of administrative data sources, including combinations of them, to estimate usual resident

CONCEPT		
All migrants arriving in the UK in local authority *j* between mid-2009 and mid-2010, residing for at least a year		

	DATA	
Source	**International Passenger Survey**	**National Insurance Number**
Period	Mid-2009 and mid-2010	Mid-2009 and mid-2010
Coverage	Sample of all arriving passengers, a subsample of which intend to stay at least 12 months in the UK (long-term international migrants)	All foreign nationals aged 16 years or over who registered for a NINo
Geographic coverage	National	National and subnational
Total number	536000 [1]	668000 [2]

	PROCESSING	
Step 1: Adjust IPS and NINo	Estimate visitor and migrant switchers and identify foreign nationals 16+ years	Adjust for arrival date where available
Step 2: Combine sources	Apportion IPS data using proportion of NINo registrations in local authority *j*	

OUTPUT		
Estimate of foreign citizens 16+ years arriving in local authority *j* in the UK and residing for at least a year, mid-year 2009 to mid-year 2010		

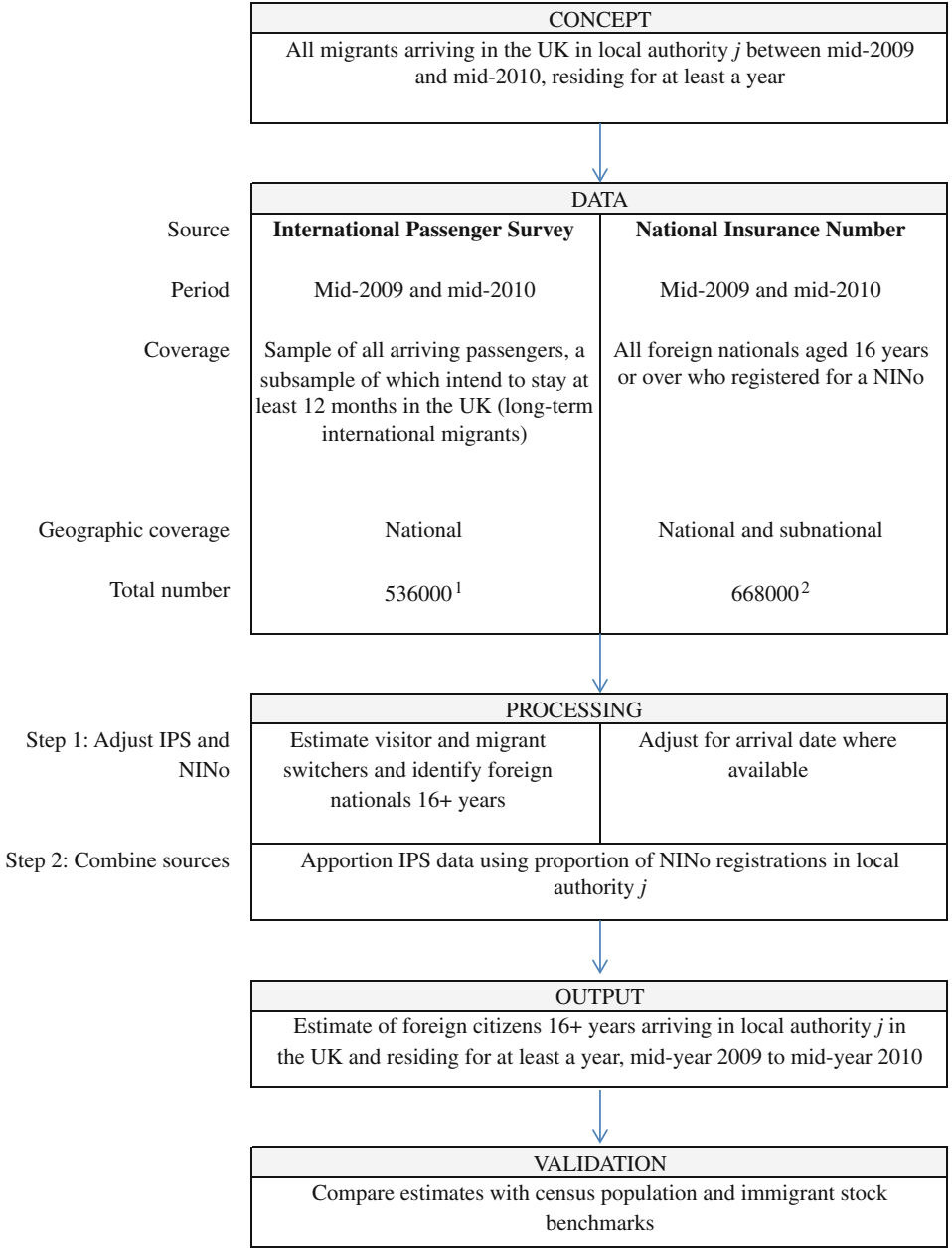| VALIDATION		
Compare estimates with census population and immigrant stock benchmarks		

Fig. 7. Application of the population statistics framework to combine International Passenger Survey (IPS) and National Insurance Number (NINo) registrations to measure international migration in the UK. Note: (1) Total immigration estimated by factoring up sample count; (2) a full count of new NINo registrations by foreign nationals.

populations as part of their 'Beyond 2011' programme (ONS 2012a). Administrative data may be used to directly or indirectly inform annual population estimates by age, sex and geography. Here, it is useful to refer to the frameworks presented in Figure 1, Figure 2 and Figure 6, as well as those presented in Zhang (2012). 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 may be conveyed visually using Venn diagrams as outlined in Figure 8 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 (or sources) in a specific area j at a specific point in time t.

To understand the relationship between a particular administrative data source and the usual resident population, consider the UK National Health Service's patient register. General Practitioners (GPs) are the first point of contact for nearly all UK National Health Service (NHS) patients (NHS 2011). Most individuals and households are registered with a GP near 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. ONS receives a list of everyone who is registered with a GP in England and Wales. This source of individual-level data has a specific administrative purpose and is not designed to specifically measure the populations.

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
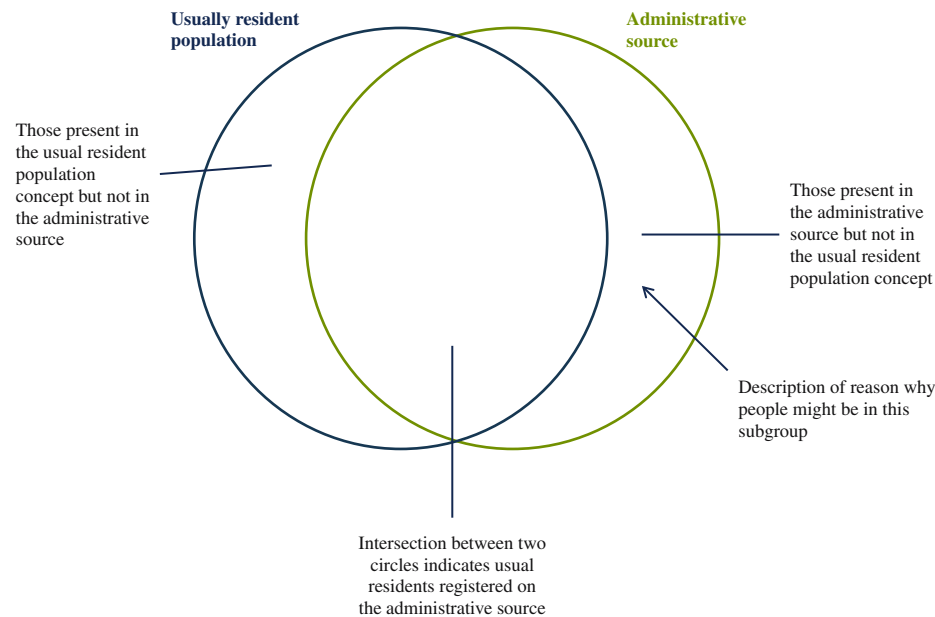

Fig. 8. Understanding the relationship between an administrative data source and the usual resident population concept

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 using their boarding school address, others using their parental address.

The coverage of the patient register is all persons registered with a GP in England and Wales. The population covered by the patient register includes the following subgroups not present in the usual resident population:

- Persons staying in the country for fewer than twelve months (*short-term migrants*);
- Patients not removed from the register because of unregistered deaths or when they have moved out of an area and not deregistered from the NHS (*list inflation or registration lag for outmigrants or emigrants*);
- Persons issued with new duplicate NHS numbers, for example when away from home (*multiple NHS numbers*).

Conversely, the usual resident population has subpopulations not recorded in the patient register:

- Persons who have failed to register at their destination after an inmigration or immigration, more likely among young adult males (*nonregistration or registration lag for inmigrants or immigrants*);
- Patients removed from the patient register when they have not moved (*erroneous list cleaning*);
- Military personnel treated by their own health services (*armed forces*); and
- Inmates of prisons treated by the prison medical service (*prisoners*).

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. Note that the NHS Patient Register is not used directly as a population estimate. Instead, patients are matched between two registers one year apart to count patient transitions. These are converted to movements before being used in the demographic accounting equation that produces the new end of internal population estimate.

Another important administrative data source is the specific social security and revenue information held within the Department of Work and Pensions (DWP), referred to as the Customer Information System (CIS) (ONS 2013d). The CIS contains a record for all individuals that have been issued with a NINo. For the DWP, the primary purpose of the CIS is to store basic identifying information, such as name, address and date of birth. The extract of the CIS provided to ONS on the 2011 Census day (27 March) contained approximately 96 million records. In Figure 9, a Venn diagram is presented showing the differences between the usual resident population with a NINo and the CIS records.

The NHS patient register provides good coverage of the population and represents the timeliest source of administrative data currently available. The CIS also provides good coverage of the population but includes a large number of people who are no longer resident in the UK. Additional administrative sources include the English and Welsh School Censuses and Higher Education Statistics Agency data (ONS 2013e), which
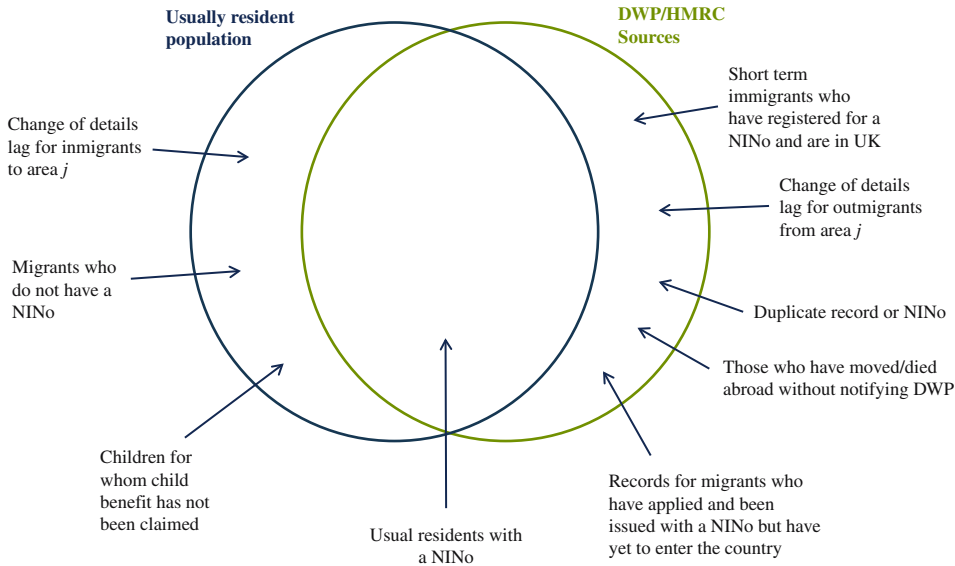
Usually resident population

DWP/HMRC Sources

Short term immigrants who have registered for a NINo and are in UK

Change of details lag for inmigrants to area *j*

Change of details lag for outmigrants from area *j*

Migrants who do not have a NINo

Duplicate record or NINo

Those who have moved/died abroad without notifying DWP

Children for whom child benefit has not been claimed

Usual residents with a NINo

Records for migrants who have applied and been issued with a NINo but have yet to enter the country

Fig. 9.   *Relating the CIS to the usual resident population in area j at time t*

provide comprehensive information but only for subsections of the population. In combination, all these sources and others have the potential to provide comprehensive and timely information about population stocks and distributions in the UK. In order to produce these estimates, ONS needs to incorporate 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 captures the concept of usual residence. Work to achieve these combined estimates has begun (ONS 2013a). One illustration of this approach is summarised in Figure 10 below within the context of the population statistics framework. For each data set an attempt to remove those individuals not resident needs to be made. It is also necessary to adjust for those who may be recorded in the wrong place. The coverage, or representation error, within all the sources is the product of a complex web of issues that needs to be disentangled. For areas where populations are relatively stable the issues have a less significant impact than for those areas experiencing high levels of population turnover. Note that national statistics agencies, including ONS, are also interested in other populations, such as present population, short-term population and visitors. Which concept is used depends on the purpose to which the population statistics will be put.

ONS is undertaking the Beyond 2011 programme to explore options for producing population and sociodemographic statistics in future. Similar work is being undertaken by National Records for Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA). Some of the options involve using administrative registers to estimate the full population while large household surveys are used to assess coverage and to add the attribute detail delivered by the traditional census (ONS 2013f). These new methods have the potential to deliver population statistics more frequently, though the precision may be lower for small-area population statistics than for large-area population statistics.

CONCEPT		
Usually resident population of local authority *j*		

DATA			
Source	**Patient register**	**Customer Information System**	**Higher Education Statistics Agency**
Date	31 July 2011	31 March 2011	2010–2011 Academic year
Coverage	Everyone currently registered with a General Practitioner in England and Wales	Everyone who has ever been registered for a NINo	All students currently registered at government-funded higher education institutions
Geographic coverage	Address	Address on date of extract	Domicile and term-time addresses
Total number	58,509,000	96,450,000	2,814,000

PROCESSING	
Step 1	Adjust sources to match required concept as closely as possible
Step 2	Anonymously link data sources
Step 3	Apply decision rules to include and exclude records
Step 4	Apply aggregate adjustments

OUTPUT
Estimate of usually resident population living in local authority *j* in 2011

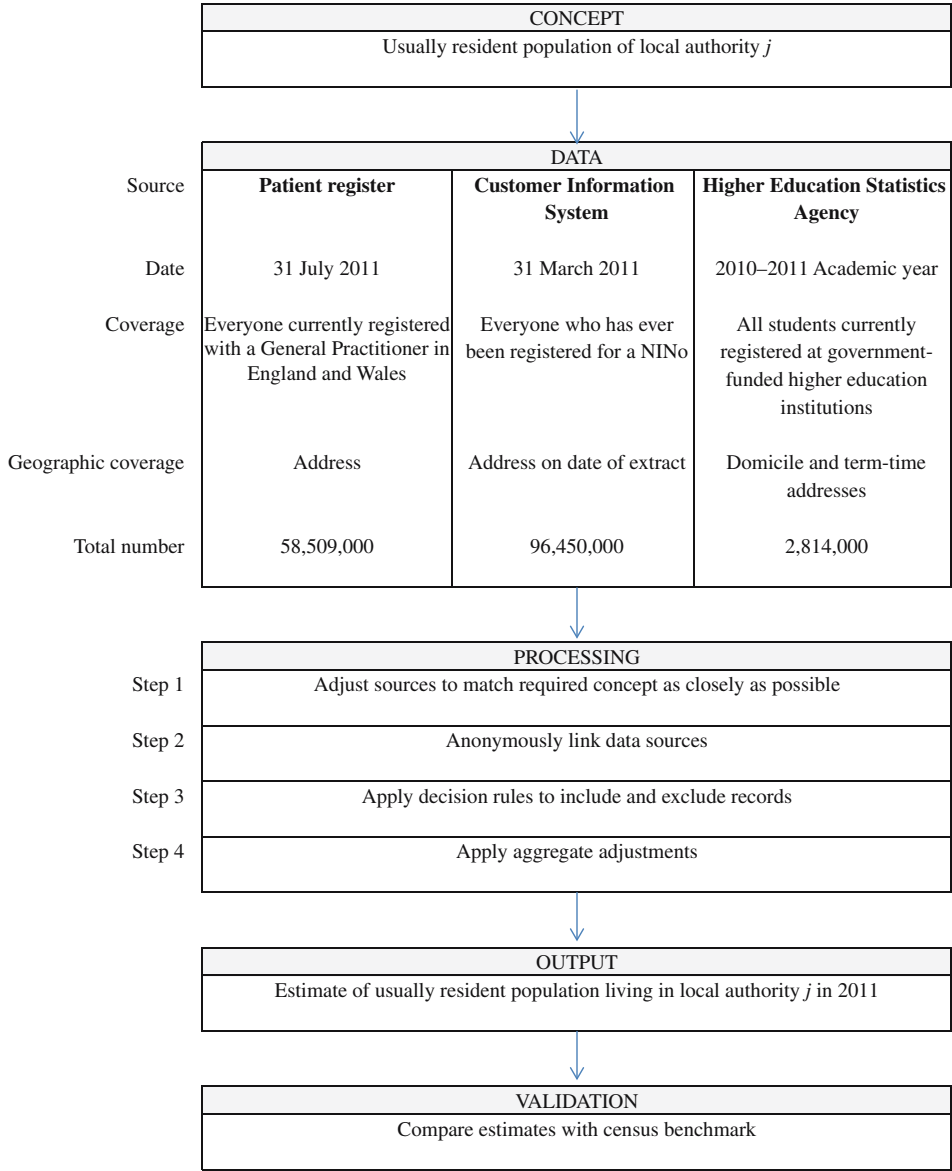| VALIDATION |
|---|
| Compare estimates with census benchmark |

Fig. 10. Application of the population statistics framework to combine administrative data sources to measure the usual resident population in the UK (ONS 2012d, ONS2013a, ONS2013d, ONS 2013e)

5.   Conclusion

The production of population 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 article we have presented central frameworks for developing and improving population estimates. The key elements are user requirement, concepts, data, processing and outputs. User requirements determine the population statistics of interest. Concepts are required to

understand and define the population statistics required 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 certain groups of the population, processing is required in order to produce the final statistical output.

The frameworks for official statistics, population statistics, demographic accounting and survey error presented in this article are 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. There are many requirements and types of population statistics but there are rarely single sources of information that cover all of these and match the concepts exactly. 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 article provides a better understanding of the relationships between the population concepts required by users and the data available to measure them. The frameworks presented in this article should also provide a guide for any country considering alternative data or methods for producing population and migration statistics. We focused on the UK and the work being carried out in the Office for National Statistics because of the transition that is occurring to incorporate and make better use of administrative data to estimate local populations. In particular, these frameworks have been used to guide the Beyond 2011 project that assessed administrative data quality, including coverage from a variety of sources, and have provided a single reference point for both users and providers.

6. References

Bell, M., M. Blake, P. Boyle, O. Duke-Williams, P. Rees, J. Stillwell, and G. Hugo. 2002. "Cross-National Comparison of Internal Migration: Issues and Measures." *Journal of the Royal Statistical Society Series A* 165: 435–464. Doi: http://dx.doi.org/10.1111/1467-985X.t01-1-00247.

Boden, P. and P. Rees. 2009. "Using Administrative Data to Improve the Estimation of Immigration to Local Areas in England." *Journal of the Royal Statistical Society Series A* 173: 707–731. Doi: http://dx.doi.org/10.1111/j.1467-985X.2009.00637.x.

Boyle, P. and D. Dorling. 2004. "The 2001 UK Census: Remarkable Resource or Bygone Legacy of the 'Pencil and Paper Era'?" *Area* 36: 101–110. Doi: http://dx.doi.org/10.1111/j.0004-0894.2004.00207.x.

Bycroft, C. 2011. *Social and Population Statistics Architecture for New Zealand.* Wellington: Statistics New Zealand.

Groves, R., F. Fowler, M. Couper, J. Lepkowski, E. Singer, and R. Tourangeau. 2009. *Survey Methodology*, 2nd ed. New York: Wiley.

Kupiszewski, M. and D. Kupiszewska. 2008. "International Migration Component in Population Dynamics Models." In *International Migration in Europe: Data, Models and Estimates*, edited by J. Raymerand and F. Willekens, 309–327. Chichester: Wiley.

Laux, R. 2002. "Review of the Framework for Labour Market Statistics." *Labour Market Trends*: 485–492. Available at: http://www.ons.gov.uk/ons/rel/lms/labour-market-trends–discontinued-/volume-110–no–9/review-of-the-framework-for-labour-market-statistics.pdf

NHS 2011. *GP Choice: Choosing a GP*. Available at: http://www.nhs.uk/choiceintheNHS/Yourchoices/GPchoice/Pages/ChoosingaGP.aspx (accessed 7 August 2012).

ONS. 2011. *Improved Methodology for Estimating Immigration to Local Authorities (LAs) in England and Wales. Office for National Statistics*. Available at: http://www.ons.gov.uk/ons/guide-method/method-quality/imps/improvements-to-local-authority-immigration-estimates/index.html (accessed 8 October 2015).

ONS. 2012a. *Beyond 2011. Office for National Statistics*. Available at: http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes—projects/beyond-2011/index.html (accessed 7 August 2012).

ONS. 2012b. *Uncertainty in Local Authority Mid year Population Estimates*. Available at: http://www.ons.gov.uk/ons/guide-method/method-quality/imps/latest-news/uncertainty-in-la-mypes/index.html (accessed 7 August 2012).

ONS. 2012c. *Estimating Internal Migration: Customer Guidance Notes*. Available at: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/internal-migration-methodology/estimating-internal-migration-customer-guidance-notes—november-2012.doc (accessed 23 August 2013).

ONS. 2012d. *Beyond 2011: Administrative Data Sources Report: NHS Patient Register*. Available at: http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes—projects/beyond-2011/news/reports-and-publications/sources/beyond-2011–administrative-data-sources-report–nhs-patient-register.pdf (accessed 23 August 2013).

ONS. 2013a. *Beyond 2011: Producing Population Estimates Using Administrative Data: In Practice*. Available at: http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes—projects/beyond-2011/news/reports-and-publications/beyond-2011-producing-population-estimates-m7.pdf (accessed 23 August 2013).

ONS. 2013b. *Beyond 2011: Reports and Publications*. Available at: http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes—projects/beyond-2011/news/reports-and-publications/index.html (accessed 23 August 2013).

ONS. 2013c. *Census 2011: Coverage Assessment and Adjustment Methods*. Available at: http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/methods/coverage-assessment-and-adjustment-methods/index.html (accessed 23 August 2013).

ONS. 2013d. *Beyond 2011: S5 Administrative Data Sources Report: Department for Work and Pensions (DWP) and Her Majesty's Revenue and Customs (HMRC) Benefit and Revenue Information (CIS) and Lifetime Labour Market Database (L2)*. Available at: http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes—projects/beyond-2011/news/reports-and-publications/sources/beyond-2011-administrative-data-sources-report—dwp-and-hmrc-cis-and-l2-combined–s5-.pdf (accessed 23 August 2013).

ONS. 2013e. *Beyond 2011: Administrative Data Sources Report: Higher Education Statistics Agency: Student Record*. Available at: http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes—projects/beyond-2011/news/reports-and-publications/sources/beyond-2011–administrative-data-sources-report–higher-education-statistics-agency-student-record–s4-.pdf (accessed 23 August 2013).

ONS. 2013f. *Beyond 2011: Options Explained 2*. Available at: http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes—projects/beyond-2011/news/reports-and-publications/methods-and-policies/beyond-2011-options-explained-2.pdf (accessed 23 August 2013).

ONS. 2013g. *1.02 IPS Margins of Error, 1975–2012. Table 1.02 Long-Term International Migration, Estimates from International Passenger Survey, Annual Data 2010, Contacts, Estimates, 95% Confidence Intervals and Ranges for the 95% Confidence Intervals for Actual Length of Stay*. Available at: http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-346438 (accessed 8 October 2015).

Poulain, M., N. Perrin, and A. Singleton. 2006. *THESIM: Towards Harmonised European Statistics on International Migration*. Louvaine: UCL Presses Universitaires.

Raymer, J. and P.W.F. Smith. 2010. "Editorial: Modelling Migration Flows." *Journal of the Royal Statistical Society, Series A* 173: 703–705. Doi: http://dx.doi.org/10.1111/j.1467-985X.2010.00660.x.

Raymer, J., P.W.F. Smith, and C. Giulietti. 2011. "Combining Census and Registration Data to Analyse Ethnic Migration Patterns in England from 1991 to 2007." *Population, Space and Place* 17: 73–88. Doi: http://dx.doi.org/10.1002/psp.565.

Rees, P. 1985. "Does It Really Matter Which Migration Data You Use in a Population Model?" In *Contemporary Migration Studies*, edited by P. White and G. van der Knaap, 55–77. Norwich: GeoBooks.

Rogers, A. 1990. "Requiem for the Net Migrant." *Geographical Analysis* 22: 283–300. Doi: http://dx.doi.org/10.1111/j.1538-4632.1990.tb00212.x.

Stone, R. 1971. "An Integrated System of Demographic, Manpower and Social Statistics and its Links with the System of National Economic Accounts." *Sankhyā: The Indian Journal of Statistics* 33: 1–184.

Statistics New Zealand. 2011. *Population Domain Plan 2012. Draft for Public Consultation*. Wellington: Statistics New Zealand. Available at: http://www.stats.govt.nz/~/media/Statistics/browse-categories/population/population-domain-plan/pop-dom-plan-draft-2012.pdf (accessed 8 October 2015).

United Nations. 1975. *Towards a System of Social and Demographic Statistics*. Department of Economic and Social Affairs, Studies in Methods, Series F, No. 18. Publication ST/ESA/STAT/SER.F/18. New York: United Nations. Available at: http://unstats.un.org/unsd/publication/SeriesF/SeriesF_18E.pdf (accessed 8 October 2015).

United Nations. 2007. *Principles and Recommendations for Population and Housing Censuses, Revision 2*, Statistical papers Series M. No. 67/Rev.2, Statistics Division, Department of Economic and Social Affairs, United Nations. New York: United Nations. Available at: http://unstats.un.org/unsd/demographic/sources/census/docs/P&R_%20Rev2.pdf (accessed 8 October 2015).

Zhang, L.-C. 2011. "Topics of Statistical Theory for Register-Based Statistics and Data Integration." *Statistica Neerlandica* 66: 41–63. Doi: http://dx.doi.org/10.1111/j.1467-9574.2011.00508.x.