
Note: These research outputs are NOT official statistics on the population

October 2015
## Contents

Executive Summary ........................................................................................................ 4  
1.0 Background ......................................................................................................... 7  
2.0 Introduction ......................................................................................................... 8  
2.1 Future releases .................................................................................................... 8  
3.0 Methodology ........................................................................................................ 9  
3.1 Administrative data used by ONS in producing population statistics ............... 9  
3.1.1 Data used in the official mid-year population estimates ............................. 10  
3.1.2 Data used in the research outputs in this paper ........................................... 10  
3.2 Methods for estimating the population ............................................................... 11  
3.2.1 Official mid-year population estimates methodology ................................. 11  
3.2.2 Beyond 2021 research outputs methodology ............................................... 11  
3.3 Development of SPD rules ................................................................................ 11  
3.4 SDP V1.0 as a stable research outputs series .................................................... 14  
3.5 Use of armed forces data in research outputs .................................................... 16  
3.6 Future developments of the SPD ....................................................................... 17  
3.7 Usual residence definition .................................................................................. 18  
4.0 SPD V1.0 Performance at National level .......................................................... 19  
4.1 Coverage of administrative data ...................................................................... 19  
4.2 The census and official mid-year population estimates as comparators .......... 20  
4.3 SPD V1.0 Comparisons to census in 2011 ....................................................... 21  
4.4 SPD V1.0 Comparison to 2013 and 2014 official mid-year population estimates .. 23  
4.5 Comparisons between SPD V1.0 and official estimates for Wales ............... 24  
5.0 SPD V1.0 analysis at local authority level ......................................................... 27  
5.1 Difference between SPDs and official estimates in 2011 ............................... 28  
5.2 Performance against P1 and P3 quality standards: 2011, 2013 and 2014 ....... 32  

Note: These research outputs are NOT official statistics on the population
Executive Summary

One of the major objectives of the 2021 Census Transformation Programme is to continue with research into future methods for producing statistics about the size of the population for England and Wales and its characteristics. This research will provide evidence to enable a decision about the future provision of population statistics after 2021. This objective is a continuation of work undertaken as part of the Beyond 2011 Programme, which culminated in the National Statistician’s recommendation in 2014 to deliver a predominantly online census in 2021, and to continue research into a future method based on administrative data and surveys beyond the 2021 Census.

The research outputs accompanying this paper are the first release of estimates of the size of the population that have been produced by pseudonymously linking multiple administrative datasets to produce a Statistical Population Dataset (SPD). The administrative datasets used to construct the SPD include the National Health Service (NHS) Patient Register, Department for Work and Pensions (DWP) Customer Information System, and data from the Higher Education Statistics Agency (HESA). SPD population estimates have been produced for each local authority, by five-year age groups and sex for 2011, 2013 and 2014. These estimates have been compared to official Census Estimates in 2011 and official Mid-year Population Estimates (MYEs) in 2013 and 2014.

To note that not all data sources were available for supply to ONS for 2012 over the reference period. Unavailable datasets for 2012 will not be supplied in future and so research outputs for 2012 will be permanently omitted from this series.

The aim of these research outputs is to:

- update users on our progress with administrative data and to seek feedback with the aim of improving our methods, and
- help the process of working with data suppliers to improve data quality

We will aim to develop them in both method and content, in the run up to the 2021 Census, and ultimately compare them with results from the 2021 Census. We will also expand on the range of topics and granularity published each year.

It is important to note that these outputs are NOT official statistics on the population. Rather they are published here as outputs from our continued research into a different methodology to that currently used in the production of population statistics. It is important that the information and research presented in this paper is read alongside the outputs to aid interpretation and avoid misunderstanding.

The results presented in this paper show some promise for this methodology. However it is clear that additional administrative data (in particular, activity data) will be required. Further, as we noted during the research carried out in the Beyond 2011 Programme, administrative data alone is unlikely to produce population statistics that are robust and of the necessary quality. A Population Coverage Survey may be needed to help assess and adjust for

---

1 Pseudonymisation is a procedure by which identifying fields (i.e. names, dates of birth and addresses) within a data record are replaced by one or more artificial identifiers to protect the privacy of individuals.

2 ‘Activity’, can be defined as an individual interacting with an administrative system, for example being in contact with tax or benefit authorities, attending hospital appointments or updating information on government systems in some other way.
coverage error in the administrative data. There are currently no plans to collect any new, regular surveys (such as a Population Coverage Survey) prior to 2021. However we will also continue our research into the development of outputs that include a coverage adjustment and where possible we will use existing survey data to do this.

The main findings presented in this paper include the following:

- SPD population estimates at national level are slightly lower than official population estimates. The SPD estimate for England and Wales was 0.93% lower than the census estimate in 2011, 0.54% lower than the MYE in 2013, and 0.85% lower than the MYE in 2014 (section 4)

- Differences between SPD population estimates and official estimates vary by five-year age group and sex. National SPD estimates for children and young adults aged between 5 and 19 are consistently lower than official estimates in 2011, 2013 and 2014. SPD estimates are consistently higher than official estimates for males aged between 35 and 55 in all three years (section 4)

- For 2011, 94% of local authorities have SPD total population estimates within +/-3.8%\(^3\) of the census estimate. This declined slightly in the comparison to MYEs in 2013 (93%), and slightly further in the comparison to MYEs in 2014 (90%) (section 5)

- The SPD method generally estimates similar levels of population change as the MYEs between 2011 and 2014 (section 5)

- A series of case studies are presented in this paper to highlight some of the quality issues associated with the use of administrative data for producing population statistics from an SPD. Examples include list cleaning in the GP Patient Register and failures in the deregistration of people when leaving a geographical area (section 6)

The research presented in this paper is based on a stable methodology that will remain constant for future releases. This will allow us to produce a consistent time series to understand how the method performs over time, and on which users can provide feedback. However we recognise the need for continued development of the methodology, so future releases will also include research based on the latest methodology which will allow us to take on board feedback that we receive from users, new data that becomes available, and alternative techniques to produce population statistics.

We also aim to expand the range of topics in future, depending on the availability of data and its quality. Possible research we are exploring include qualifications, personal or household income, and ethnicity, as well as on the number and the size and composition of households.

As a guide to the remaining sections of this paper,

- Sections 1 and 2 provide some background and an introduction

- Section 3 describes the methodology used to produce the research outputs, with consideration for how the SPD methodology differs from the existing method used for mid-year estimates (MYE)

---

\(^3\) 3.8\% is equivalent to approximately 6,000 people in an average sized local authority of around 160,000 people

Note: These research outputs are NOT official statistics on the population
• Section 4 provides a summary of national performance of the SPD and outlines some of the coverage issues associated with the use of administrative data

• Section 5 presents a more detailed analysis of SPD performance at local authority level by sex and five-year age groups

• Section 6 contains a series of detailed case studies that highlight some of the particular quality issues relating to the accuracy of administrative data, including timeliness of updates and operational processes

This is accompanied by the research output tables themselves which can be found HERE and a template for users to provide feedback HERE.

Note: These research outputs are NOT official statistics on the population
1.0 Background

The Census Transformation Programme was established in January 2015 to take forward the National Statistician’s recommendation on the future of the census and population statistics made in March 2014. The recommendation to make the best use of all available data in the production of population statistics is reflected in the three major objectives of the Programme, to deliver:

- a predominantly online census of all households and communal establishments with special care taken to support those who are unable to complete the census online
- improved and expanded population statistics through increased use of administrative data and surveys, and
- evidence to enable a decision about the future provision of population statistics after 2021

The Government’s response to the National Statistician’s recommendation on the future of the census and population statistics made reference to the following:

“The Government welcomes the recommendation for a predominantly online census in 2021 supplemented by further use of administrative and survey data.

Government recognises the value of the census and its history as a bedrock of statistical infrastructure. The census provides information on the population that is of fundamental importance to society....

Our ambition is that censuses after 2021 will be conducted using other sources of data and providing more timely statistical information .... dependent on the dual running sufficiently validating the perceived feasibility of that approach.”

Building on the work carried out during the Beyond 2011 Programme, this work involves research into the potential use of administrative data and surveys to produce population, household and characteristic information currently provided by a census. Key to this will be assessment criteria that will need to be met to provide evidence to move away from the ten yearly census approach after 2021, including comparing administrative and survey based estimates with those from the 2021 Census. It will also take forward the recommendations that were outlined in the Beyond 2011: Independent Review of Methodology led by Professor Chris Skinner.

An annual assessment showing progress will be published, starting in spring 2016. Supporting these annual assessments, and also the second objective to improve the range of population statistics through increased use of administrative data and surveys, will be the production of an annual research update including administrative data based research outputs from autumn 2015. This will culminate in a recommendation about the future provision of population statistics after 2021.

Note: These research outputs are NOT official statistics on the population
2.0 Introduction

This paper is the first of a series of annual research updates on progress made with administrative data to support the objectives of the Programme. Although primarily focussed on developing the evidence to support the future recommendation on population statistics and censuses after 2021, we will also report on progress towards the acquisition of administrative data for statistical use by the Programme to deliver new outputs from the 2021 Census. Progress so far is described in the Administrative data update.

This is the first release in an annual series of research outputs. This year, these include administrative data based estimates of the population in each local authority for each five-year age group and sex of the population for years 2011, 2013 and 2014. The methodology to produce these has been further developed since that presented during the Beyond 2011 Programme and the paper describes these.

To note that not all data sources were available for supply to ONS for the 2012 reference period. Unavailable datasets for 2012 will not be supplied in future and so outputs for 2012 will be permanently omitted from this series.

It is important to note that these outputs are NOT official statistics on the population. Rather they are published here as outputs from our continued research into a different methodology to that currently used in the production of population statistics. It is important that the information and research presented in this paper is read alongside the outputs to aid interpretation and avoid misunderstanding.

We are using the existing official population statistics to assess the quality of the outputs presented here. Whilst there are currently no plans to put this methodology into production prior to 2021, ONS are using these findings to understand any implications for the current methodology, and will introduce any improvements where appropriate.

We are keen to get feedback on the research outputs and the methodology and how they might be improved. Users are encouraged to assess them and send feedback via the template available HERE, or by contacting Beyond.2021.Research.and.Design@ons.gov.uk. We will also be hosting a range of meetings throughout December and January 2016 as a route for getting user feedback on these outputs. Such feedback will feed into an assessment to be published on an annual basis in the spring (see section 1).

2.1 Future releases

The outputs presented in this paper are based on a stable methodology that will remain constant for future releases. This series has been labelled the ‘SPD V1.0’ series and will be replicated each year. This will allow us to produce a consistent time series to understand how the method performs over time, and for users to provide feedback on. However we recognise the need for continued development of the methodology, so future releases will also include outputs based on future developments of the methodology which will allow us to take on board feedback that we receive from users, new data that becomes available, and alternative techniques to produce population statistics. A new series will be introduced in 2016 to include these developments which will show improvements to the methodology.

Additionally, subject to satisfactory data access and quality, we also aim to release population statistics at more detailed levels, and outputs about other topics or characteristics of the population (other than age and sex) such as qualifications, personal or household income, and ethnicity, as well as on the number and the size and composition of households. We also plan to present our research into the development of a methodology to estimate and adjust for coverage error in the SPD.

Note: These research outputs are NOT official statistics on the population
3.0 Methodology

The set of research outputs accompanying this paper have been produced from a Statistical Population Dataset (SPD), which has been constructed by pseudonymously linking records across multiple administrative data sources and applying a set of ‘inclusion and distribution rules’. Our policy on safeguarding data when linking multiple administrative datasets is set out in our previously published Beyond 2011 paper, Safeguarding Data.

The data sources used include the NHS Patient Register, DWP’s Customer Information System, and data from the Higher Education Statistics Agency. Links to an ONS assessment of the quality and potential for using these administrative datasets for population estimates are provided below:

- NHS Patient Register Beyond 2011 Source Report
- Customer Information System Beyond 2011 Source Report

Aggregate statistics supplied by the Ministry of Defence (MoD) for home and foreign armed forces personnel are also included in the SPD estimates. A glossary has been provided at Annex A to summarise the content of data used in this research, as well as explanations for some of the technical concepts referred to in this paper.

Population estimates\(^4\) by five-year age group and sex at local authority level have been derived from the SPD and are the basis for comparison against official ONS mid-year population estimates (MYEs).

While there is overlap in some of the administrative data sources used in the production of official MYEs, there are some important differences relating to the methodology for constructing an SPD and MYEs. This section summarises those differences and outlines how the SPD rules have been updated since the Beyond 2011 programme to introduce the new research outputs series starting in 2015. This series will be published annually in the run up to the 2021 Census, with the intention of continually improving the method for producing population estimates from administrative data.

3.1 Administrative data used by ONS in producing population statistics

This section provides information about administrative sources currently used for official population statistics as well as administrative sources that have the potential to be used in the future.

\(^4\) To note: throughout this paper, we refer to the data presented as ‘population estimates’. These are estimates directly produced from the linked administrative data, without adjusting for biases in the relating to coverage of the administrative data. Statistical estimates of the population with associated variance calculations may be produced in future with combined use of a Population Coverage Survey.
3.1.1 Data used in the official mid-year population estimates

The mid-year population estimates use a combination of sources to produce estimates down to local authority level, including administrative data and surveys. These are summarised below:

- Birth and death registrations from the General Register Office;
- International Passenger Survey data, Labour Force Survey data and administrative data sources used to estimate international migration moves, and Home Office data to estimate the number of asylum seekers and their dependants;
- National Health Service Central Register, the GP Patient Register Data System and Higher Education Statistics Agency data used to estimate internal migration moves;
- Defence Analytical Services & Advice and United States Air Force data used to estimate the change in the home and foreign armed forces population;
- Ministry of Justice data used to estimate the number of prisoners.
- ONS census estimates used to rebase estimates for each decennial series.

3.1.2 Data used in the research outputs in this paper

Three core administrative datasets are used in the current method for linking records and constructing an SPD to derive population estimates:

- The Patient Register (PR), which contains a list of all patients who are registered with an NHS GP in England and Wales;
- The Customer Information System (CIS), which contains a list of all people who have a National insurance Number;
- The Higher Education Statistics Agency data (HESA), which contains a list of students who are registered on to a Higher Education course in England and Wales.

As described in section 3.0, aggregate statistics supplied by the Ministry of Defence (MoD) for home and foreign armed forces personnel are also included in the SPD estimates by adding them to the counts produced by linking the above sources. We are currently in the process of working with other departments across government to explore the potential of accessing additional datasets to support our work in developing research outputs. An update on our progress in pursuing administrative data for use in the Census Transformation Programme has been provided to accompany this research paper HERE.
3.2 Methods for estimating the population

3.2.1 Official mid-year population estimates methodology

MYEs are produced from a cohort-component method which uses information from administrative data sources and surveys to update the population base annually. In census year, the series is rebased with the official census estimates. In order to produce estimates during the intercensal period, MYEs rely on measuring flows of the population across four components: fertility, mortality, internal migration and international migration. Further information on the methods involved in the production of the MYEs at local authority level can be found on the ONS website. The most difficult challenge of producing MYEs is estimating migration. This results in a reduction in the accuracy of estimating the population during the intercensal period that is observable, and corrected following census rebase. A report outlining ONS’ plans to review the methodology relating to estimating migration in MYEs can be found in the 2014 user update papers for internal migration and international migration.

3.2.2 Beyond 2021 research outputs methodology

The Beyond 2021 research outputs have been produced by matching pseudonymised individual records across the administrative data sources listed above. The methods used to undertake this matching are outlined elsewhere in our Matching Methodology Paper. The data is pulled together into a single dataset and estimates produced from a set of inclusion, exclusion and distribution rules. Unlike the MYEs, the SPD is a stock based approach at a particular point in time. It is therefore dependent on the administrative data having good coverage of the usual resident population (see section 3.7 below), and is more vulnerable to changes in data over time. However, assuming the operational processes for collecting data are stable, it generally has the advantage of producing estimates of a similar level of accuracy each year.

3.3 Development of SPD rules

During the Beyond 2011 Programme a set of SPDs were constructed in a stepwise manner from administrative data that was supplied for 2011. A total of twelve SPDs were created over the period, the first six of which were the focus of our research published in Estimating the Population: In Practice. Different rules were applied in the construction of each SPD based on consideration for how the different administrative sources capture and record the targeted usual resident population.

Table 1 shows the performance of SPDs 1 to 6 compared with official 2011 Census estimates. The P1 and P3\(^5\) quality standards included in table 1 are outlined in the Beyond 2011 Final Options Report, and are summarised in section 4.2. These quality standards were developed as part of the Beyond 2011 evaluation criteria for making the recommendation about the census in 2021. While these quality standards are still being used to evaluate SPD performance in this year’s release, they may change in subsequent years following any future review.

---

\(^5\) Paper O4 described three different quality standards. For simplicity, we mainly use the strictest quality standard P1, in this paper as the basis of comparison.
A full description of the rules applied and the performance of SPDs 1-6 can be found in our previously published paper *Estimating the Population: In Practice*.

**Table 1, Percentage of SPD estimates for local authority areas in England and Wales meeting quality standards P1 and P3**

<table>
<thead>
<tr>
<th>Data source(s) used</th>
<th>Rule to distribute records if LA is not the same across sources</th>
<th>P1 Quality standard</th>
<th>P3 Quality standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPD PR CIS HESA</td>
<td>Exclude records where LA is not the same on both sources</td>
<td>Weight of 0.5 applied to each LA*</td>
<td>Weight applied based on evidence found on census*</td>
</tr>
<tr>
<td>1</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>✓ ✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>✓ ✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>✓ ✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>✓ ✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

*Except in instances where an individual is found on HESA as well as PR and CIS (SPDs 5 and 6). In these circumstances the whole record is allocated to term-time address from the HESA record

Table 1 sets out the rules used to determine which records are to be included in the population estimate, as well as rules for distributing records to geographic location. Due to time lags in the process of registration and updating of administrative records, it is often the case that there is conflict regarding the address of an individual linked across administrative sources. In such cases, the SPD rules rely on either a deterministic assignment or the apportioning of weights for estimating records in geographic locations.

Although SPD 6 is the closest to meeting the P1 quality standard, its construction relies on the 2011 Census to produce distribution weights for apportioning records to geographic locations. There is uncertainty as to whether such an approach could be reliably informed by a PCS that has been undertaken on a sample basis in the absence of a census. For this reason SPD 5 was evaluated to be the best performing SPD on the basis that it is replicable, independent of census or survey data, and performed comparatively well against the quality standards. For SPD 5, 90.8% of the total population estimates for local authorities were within +/- 3.8% of 2011 Census estimates.

The following rules were used to construct SPD 5 from the administrative data sources listed previously:

- If an individual has been linked between PR and CIS, and their address is in the same LA, include the linked record and apply a weight of 1.0
- If an individual has been linked on PR and CIS, and their address is in a different LA, include both records but apply a weight of 0.5 to each LA
If an individual has been linked between PR, CIS and HESA, then assign a weight of 1.0 to the LA area recorded on HESA.

A further six SPDs were introduced towards the end of the Beyond 2011 Programme, most of which extended on rules used to construct SPD 5. These SPDs were the focus of our supplement to the Statistical Research Update Paper published in January 2014. Table 2 below shows the performance of SPDs 7 to 12 compared to 2011 Census estimates.

**Table 2, Overview of SPD performance (in percentage) against quality standards P1 and P3**

<table>
<thead>
<tr>
<th>Data sources used</th>
<th>Rule to include records on the SPD</th>
<th>Rule to distribute records if LA is not the same across sources*</th>
<th>P1 Quality standard</th>
<th>P3 Quality standard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exclude records where OA is not</td>
<td>Inclusion on SPD where HESA links to either PR or CIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>the same on both sources</td>
<td>Inclusion of all records on PR for 0 to 4 year olds</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weight of 0.5 applied to each LA*</td>
<td>Distribute all records to PR LA*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distribute all records to CIS LA*</td>
<td>Distribute records to LA recorded on School Census</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPD</td>
<td></td>
<td></td>
<td>Within +/-3.8%</td>
<td>Within +/-8.5%</td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
<td>✓</td>
<td>13.2</td>
<td>67.8</td>
</tr>
<tr>
<td>8</td>
<td>✓</td>
<td>✓</td>
<td>92.5</td>
<td>98.9</td>
</tr>
<tr>
<td>9</td>
<td>✓</td>
<td>✓</td>
<td>92.8</td>
<td>98.9</td>
</tr>
<tr>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>92</td>
<td>98.9</td>
</tr>
<tr>
<td>11</td>
<td>✓</td>
<td>✓</td>
<td>85.1</td>
<td>98.3</td>
</tr>
<tr>
<td>12</td>
<td>✓</td>
<td>✓</td>
<td>92.2</td>
<td>98.9</td>
</tr>
</tbody>
</table>

*Except in instances where an individual is found on HESA and PR and CIS (SPDs 7, 9, 10, 11, 12), or found on HESA and either PR or CIS (SPD 9). In these circumstances the whole record is allocated to term-time address.

SPDs 8, 9 and 12 performed better against the quality standards than SPD5. SPD8 incorporated the rule that records from HESA be included based on a single link to either PR or CIS. This attempts to address the issue of low estimates on the SPD for those aged 20 to 24 by including more students that were probably missed by the rules used to construct SPD5. SPD9 was developed to try and improve the estimate of 0 to 4 year olds on SPD 5 by including all 0 to 4 year olds on the PR regardless of whether they were present on the CIS. The impact of these two rules when applied separately as SPD8 and SPD9 is summarised in our previously published Statistical Research Update Paper. The distribution rule associated with SPD 12, where children are counted wholly in the term time location recorded on the School Census, made very little difference to the performance of the SPD for 5 to 15 year olds. In future developments we will consider the impact of including School Census records that match to either the PR or the CIS and whether it improves coverage for these age groups at the national level.

The two inclusion rules introduced in SPDs 8 and 9 have now been incorporated into a new SPD that will form the basis of a stable series for research outputs in the run up to 2021.
3.4 SPD V1.0 as a stable research output series

The research outputs published here for 2011, 2013 and 2014 are based on an SPD that combines rules across multiple SPDs tested in the Beyond 2011 Programme. Consequently the estimates produced perform better against the P1 quality standard. This SPD has been labelled SPD V1.0 and forms the basis of a stable SPD time series that will be replicated for future annual research outputs. The rules for inclusion and exclusion of records in SPD V1.0 and their justification are set out in Table 3 and Table 4.

Table 3, Inclusion rules for SPD V1.0 with their justification

<table>
<thead>
<tr>
<th>Inclusion Rule</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any record on both the PR and CIS</td>
<td>If the individual is present on both the PR and the CIS then this suggests they are resident in the country</td>
</tr>
<tr>
<td>All 0 to 4 year olds on the PR are included</td>
<td>Children only appear as a record on the CIS if a parent has registered their child to claim child benefit, therefore, the PR is a more complete source for these ages. As young children have frequent contact with the health service this source is also likely to be up to date.</td>
</tr>
<tr>
<td>HESA records if they are present on either PR or CIS or both</td>
<td>Some overseas students choose not to work whilst they are at university, whilst others are unable to work due to visa restrictions. By including those individuals on HESA if they are present on just one other source, it more accurately reflects the university student population.</td>
</tr>
</tbody>
</table>
Table 4, Exclusion rules for SPD V1.0 with their justification

<table>
<thead>
<tr>
<th>Exclusion Rule</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Records with a CIS country code other than ‘GBR’, ‘ENG’, ‘WAL’ or if country code is ‘NKN’ and there is no other geographical information.</td>
<td>A country code outside of Great Britain suggests the individual is not resident in the reference population. A ‘not known’ country code provides no evidence the individual is resident. (See table 5 for the distribution rule applied to CIS records that have a ‘GBR’ country code with an address registered in a Scottish local authority)</td>
</tr>
<tr>
<td>Records with a date of death prior to the reference date (30 June each year) provided from the CIS data</td>
<td>If there is a date of death prior to the reference date then the individual is not part of the usually resident population</td>
</tr>
<tr>
<td>Records with an FP69 flag on the PR</td>
<td>This flag indicates that there has been no contact between the GP practice and the individual in the last 15 months. An attempt at contact has been made by the GP practice but no reply has yet been received. It can’t be assumed that the individual still resides at the address. (This rule will be reviewed when constructing new SPDs for future releases as in some local authorities it may be resulting in large reductions in the population estimate. See section 6.5 for further discussion).</td>
</tr>
</tbody>
</table>

The inclusion and exclusion rules are used to identify records that make up the SPD population estimates at national level. Where there is evidence of a location conflict across the administrative sources, rules for distributing records at the local authority level are applied as presented in Table 5.
Table 5, Rules to distribute records at the local authority level with justification

<table>
<thead>
<tr>
<th>Distribution Rule</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Records with the same local authority on PR and CIS are allocated with a weight</td>
<td>This is the local authority we can assume the individual is resident at based on the evidence from the two sources.</td>
</tr>
<tr>
<td>of 1.0 (whole) to that local authority</td>
<td></td>
</tr>
<tr>
<td>Records with a different local authority on the PR and CIS are allocated with a</td>
<td>As there is no information to support allocating a whole person to one of the local authorities, a half weight is allocated to each. In other</td>
</tr>
<tr>
<td>weight of 0.5 (half) to each local authority</td>
<td>words we estimate the individual half in one LA and half in the other.</td>
</tr>
<tr>
<td>Records that have HESA information are allocated with a weight of 1.0 (whole)</td>
<td>On the basis that data on students is collected an updated by universities annually, we anticipate that term time address information on HESA is</td>
</tr>
<tr>
<td>to the local authority associated with their HESA term time address unless they</td>
<td>more reliable for students than that of PR or CIS.</td>
</tr>
<tr>
<td>have a HESA ‘remove flag’6. In the case of a HESA remove flag, we revert to the</td>
<td></td>
</tr>
<tr>
<td>previous two rules in this table.</td>
<td></td>
</tr>
<tr>
<td>Records that have a ‘GBR’ country code on the CIS with an address registered in a</td>
<td>This is likely to be the result of cross border moves between England and Wales, and Scotland. In cases that the PR is more up to date</td>
</tr>
<tr>
<td>Scottish local authority are counted with a weight of 1.0 (whole) in the local</td>
<td>(which is generally more likely than the CIS) this rule enables the correct inclusion of usual residents on the SPD.</td>
</tr>
<tr>
<td>authority of the matching PR record.</td>
<td></td>
</tr>
</tbody>
</table>

3.5 Use of armed forces data in research outputs

The SPD totals produced from linking records from the PR, CIS and HESA are also supplemented with aggregate statistics on home and foreign armed forces personnel. These four sources are combined to produce the SPD V1.0 estimate. Research in the Beyond 2011 Programme identified that there was a notably lower SPD estimate in local authorities that contain large military bases, particularly for young males of working age. The inclusion of aggregate statistics supplied by the MoD greatly improved the accuracy of SPD estimates in these areas, most likely because armed forces personnel were previously being excluded from the SPD estimate on the basis that they are not registered with NHS GPs. This approach may change in future SPD development, dependent on our improved

---

6 The HESA remove flag indicates those who are on sabbatical, dormant, visiting/exchange, have died or transferred to another institution. It indicates that the address information from HESA may not be up to date.

Note: These research outputs are NOT official statistics on the population.
understanding of how armed forces personnel, and particularly their dependents, interact with GP services. Issues relating to local authorities with a high proportion of armed forces personnel resident in the population are discussed further in section 4.1.

### 3.6 Future developments of the SPD

Alongside SPD V1.0 a second SPD series will be produced from 2016 onwards which will show developments to the methodology. This will incorporate improvements to the SPD methodology that have emerged from research over the previous year. Some areas of research that may be looked into and incorporated into this future developmental SPD include:

- Improvements in the matching methods, for example the incorporation of matches made through probabilistic methods. The current method for identifying matches between datasets in the construction SPD V1.0 relies on exact and deterministic methods (see [Matching Methodology Paper](#)). Additional use of probabilistic methods that have been developed over the last year is expected to increase the number of matches, and therefore improve the coverage of the SPD. However, there is also the possibility that probabilistic matching will increase the number of false matches, which may result in overestimating the population.

- Making use of activity data as exclusion criteria to reduce the over-coverage on the SPD. An indication of recent activity, for example economic activity relating to payment of tax or claiming benefits may be used as the basis for confirming that someone is more likely to be usually resident in the population.

- Refining the use of weights in allocating records to geographic location to improve the accuracy of where people are located based on administrative data. However this would need further evidence from administrative sources (such as activity data – see above) or surveys to make decisions regarding the most likely place of residence.

- The use of other inclusion rules, for example, perhaps using HESA data to better capture foreign students who may not appear on either the PR or CIS.

In the longer term there is also the prospect of the SPD incorporating the use of additional administrative data sources. We are currently in the process of engaging with a number of data suppliers for potential access to more data, including data from health and education agencies, DWP, HMRC, DVLA, and MoD. As part of the [feedback](#) process we are keen to hear about other sources of data and specific ideas about how we could develop the methodology for improving the SPD.

Published alongside this report is a paper that outlines the progress that we are making with acquiring administrative data to support the development of our research outputs. This paper summarises the data that we are pursuing and puts them in the context of topics are typically collected by a census, and can be found [HERE](#).
3.7 Usual residence definition

The UN definition for population estimates is based upon usual residence. This includes people who reside in England and Wales for at least 12 months, regardless of their nationality and excludes short-term migrants and visitors. This definition is central to the census and MYE methodology from design through to publishing on the basis that they aim to estimate the usual resident population at a specific point in time.

Administrative data are primarily collected for operational purposes and not designed to specifically capture usual residence. The methodology for the research outputs series has been developed with the intention of following this definition in as far as the data will allow. This has been done by selecting administrative data sources that have wide population coverage and by using the inclusion, exclusion and distribution rules presented above.

There are gaps in the coverage offered from the administrative data used to construct SPD V1.0. While aggregate statistics have been included for home and foreign armed forces personnel, there are other special population groups that will be inconsistently recorded or missing from the administrative data. Areas most affected by the absence or inconsistencies relating to special population groups will be those with significant prison populations or school borders. Further consideration is needed to improve the coverage of these groups on the SPD. In future we will be exploring the availability of specific record level data that has the potential to provide good coverage of these population groups. Further, it is unlikely that that administrative data alone will be able to provide an exact fit to the UN definition of usual residence. We will look to ‘correct’ any remaining definitional differences though the use of a coverage survey.

Note: These research outputs are NOT official statistics on the population
4.0 SPD V1.0 Performance at National Level

This section evaluates SPD V1.0 population estimates for England and Wales for the years 2011, 2013 and 2014. The evaluation is based on comparison with official estimates. Census estimates are used as the basis for comparison for 2011 estimates, MYEs as the basis for comparison in 2013 and 2014. An analysis of SPD performance at the national level provides important context for understanding some of the coverage problems observed at the local authority level which is the focus of section 5.

4.1 Coverage of administrative data

There are several reasons why differences between SPD V1.0 estimates and official population estimates may arise. These are summarised below:

**Individuals that are not usually resident appearing on administrative datasets:**
Examples include short-term migrants who intend to stay in the UK for less than 12 months, or persons who have emigrated from the country. The 2011 Census showed 195,000 short-term migrants in England and Wales and a proportion of these may appear on both the PR and the CIS and therefore will be included on the SPD. People who are no longer resident will also appear on the SPD if they have not been deregistered with their GP and failed to notify DWP or HMRC of their departure from the country. Emigrants and short term migrants that appear on both the PR and CIS will have the effect of inflating the population estimate on the SPD.

**Individuals that are usually resident but do not appear on administrative datasets:**
There are many circumstances that may result in usual residents not appearing on either the PR or the CIS. Migrant dependents that have not taken up employment or benefit claims may not have applied for a National Insurance number and will be missing from the CIS. In addition, people may not necessarily register for health services when migrating into or around the country. Non-registration of residents on the PR and CIS has the effect of producing lower estimates on the SPD. While undercoverage is typically an issue relating to census collections (due to non-response), it is adjusted for in the census estimation process (see paper *Census 2011 Assessment and Adjustment Process*). The effect of this adjustment is also carried forward in the MYEs. It should be noted that coverage problems associated with the SPD will also need to be adjusted for with a suitable method. The executive summary of this report makes reference to the possible need for a Population Coverage Survey to support the framework for producing population estimates from administrative data. For more detailed information see our previously published *Estimating the Population: In Theory* paper.

**Duplication of estimates for members of ‘special populations’:**
Census and mid-year estimates also make adjustments for special populations that are difficult to enumerate or measure in terms of population flow. An example is the home and foreign armed forces population that are added in annually at the aggregate level to MYEs. With an SPD approach the process of including aggregate armed forces statistics runs the risk of duplication in the population estimates. This would occur in circumstances where individuals included in the aggregate statistics have already been captured on the SPD. The methodology used in this year’s set of research outputs includes aggregate armed forces.
statistics on the basis that without their inclusion, noticeably low estimates were observed in the LAs containing large military bases. There is uncertainty about the extent to which double counting may occur between the administrative data and aggregate statistics on home and foreign armed forces. Total aggregate numbers for home and foreign armed forces personnel of around 157,000 in 2011 have since decreased to around 139,000 in 2014.

Reduced accuracy of MYEs during the intercensal period:
As outlined in section 2.2, the accuracy of MYEs tends to reduce during the intercensal period. As a consequence, differences observed between the SPD and the MYE may not always be indicative of coverage problems on the SPD. While the 2011 Census population estimates provide a reliable basis for comparison in 2011, there is some uncertainty about the use of MYE as the basis for evaluation in future years, particularly as the series progresses through the decade. An important part of our research in the run up to the 2021 Census is to consider alternative approaches for evaluating the quality of population estimates without reliance on census or MYEs. This will be an important aspect to the design of a statistical system built around administrative data and surveys. In the absence of any alternative approach, the MYEs remain the best comparator.

4.2 The census and official mid-year population estimates as comparators
The 2011 Census produced an estimate for the England and Wales usually resident population with a 95% confidence interval of ±0.15%. At the mid-point of the last decade the mid-year estimates have been estimated to have a precision of around ±0.22%, and around ±0.27% at the end of the decade (just before the 2011 Census). These levels of accuracy have been used to derive the national P1 and P3 quality standards against which the SPDs have been assessed.

For five-year age groups by sex, 97% of groups in the 2011 Census had a 95% confidence interval of ±0.6% or better. For mid-year estimates at the mid-point of the last decade this interval is estimated to have been ±0.8%, while at the end of the decade, it was ±1.0%. These intervals correspond to the quality standards for five-year age groups by sex (see the Beyond 2011 Final Options Report for more detail).

Table 6, Summary of national quality standards

<table>
<thead>
<tr>
<th>Quality standard</th>
<th>England and Wales</th>
<th>England and Wales – five-year age and sex groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 (maximum quality in the current system)</td>
<td>The national population estimate has a 95% confidence interval ±0.15%</td>
<td>97% of national estimates by five-year age and sex group have a 95% confidence interval of ±0.6% or better</td>
</tr>
<tr>
<td>P3 (average quality in the current system)</td>
<td>The national population estimate has a 95% confidence interval ±0.22%</td>
<td>97% of national estimates by five-year age and sex group have a 95% confidence interval of ±0.8% or better</td>
</tr>
</tbody>
</table>

Note: These research outputs are NOT official statistics on the population
4.3 SPD V1.0 Comparison to census in 2011

The 2011 Census estimated that there were 56,075,912 usual residents in England and Wales. The SPD contains 55,552,637 people, and is therefore 0.93% lower than the census population. This means that at the national level the SPD tends to produce a lower estimate of the population, which is outside the P3 quality standard.

While the SPD estimate is lower than the national aggregate level, it should be noted that this is the net effect of individuals that are missing from the SPD. The SPD could incorrectly include some records, but these are outweighed by a larger number of gross underestimate. The overall population figure conceals different levels of coverage for different groups of the population. In particular, there is substantial variation in the performance of the SPD by sex and across the range of ages. Figure 1 shows the percentage difference between the SPD and the census in 2011.

Figure 1, Difference between SPD V1.0 estimates and census estimates in 2011 by five-year age and sex group

Figure 1 shows that for most five-year age and sex groups, the difference to the census estimate is considerably larger than the interval given in the P3 quality standard. The lower SPD V1.0 estimate is prevalent for most age groups, but for males aged between 30 and 59.
there is actually a higher estimate produced on SPD V1.0. This is believed to be due to people who have left the country but have not informed the health or tax authorities of this, and so their record remains on the Patient Register, or the CIS as still living in the country. We have previously published research (see Statistical Research Update Paper) that demonstrated the effectiveness of using activity data on a 1% sample of the population. Removing records from the SPD based on the absence of activity relating to tax and benefits had the effect of reducing the overestimation of males aged 30 to 59. We are in the process of pursuing feasibility data with similar activity indicators with more complete coverage of the adult population.

The tendency of SPD V1.0 to overestimate males of this age group contrasts strongly with females, with equivalent SPD V1.0 estimates generally being about 2% lower on the SPD than estimated on the census.

Lower estimates on SPD V1.0 are most evident for children and young adults up to age 19. For 0 to 4 year olds, the update to the rules used to construct SPD V1.0 to include all records from the Patient Register has substantially improved the level of coverage for this age group compared to SPD5 (see the supplement to our previously published Statistical Research Update Paper). However, children not appearing on the CIS due to parents not claiming Child Benefit may be contributing to the low SPD V1.0 estimate that is seen for older children.

It should also be noted that SPD V1.0 produces lower estimates than the census for persons aged over 85. Further analysis is required to explore possible reasons for this difference. One possible explanation may relate to problems with matching records for persons that have recently moved into residential or nursing care homes.
4.4 SPD V1.0 Comparison to 2013 and 2014 Official Mid-year population estimates

In the years following the census, SPD V1.0 is compared to the mid-year population estimates. As shown in Figures 2 and 3, the comparison between SPD V1.0 and the mid-year estimates shows the same broad pattern to the comparison between the SPD V1.0 and census in 2011. However, a few changes have occurred.

Figure 2, Difference between SPD V1.0 estimates and mid-year estimates in 2013 by five-year age and sex group

In 2013 SPD V1.0 appears to show a slight reduction in the difference between the SPD estimate and the official estimate (MYE) for working age females. For females aged 20 to 24 and both males and females aged 25 to 29, there is actually a higher estimate on SPD V1.0 compared to the MYE. The difference between SPD V1.0 estimates and the MYE also appears to reduce for all 10 to 14 year olds and 15 to 19 year old females. Males aged 15 to 19 and 20 to 24 are exceptions to this trend, showing larger differences in 2013 in contrast to the change for females of the same age. There is also a slight increase in the differences for 0 to 4 year olds and 5 to 9 year olds.

Note: These research outputs are NOT official statistics on the population
In 2014, the SPD V1.0 estimate for working age females is slightly lower in comparison to MYEs, when compared to 2011 and 2013. For males aged 15 to 19 and 20 to 24 the increase in difference to the official estimates between 2011 and 2013 continues into 2014. The SPD V1.0 estimate for 20 to 24 year old males is now over 4% below the official estimates, compared to around 1% in 2011, so the two sets of estimates appear to be diverging over time for this group. The lower SPD V1.0 estimates for children aged 0 to 4 and 5 to 9 also increases again. For 10 to 14 year olds, the SPD estimate is getting closer to the MYE, continuing the trend seen between 2011 and 2013.

These trends could be due to changes in the administrative data affecting SPD V1.0, such as fewer parents claiming Child Benefit. The precision of the mid-year estimates is reducing however, and uncertainties are likely to magnify over time and create larger differences with SPD1.0.

4.5 Comparison between SPD V1.0 and official estimates for Wales

Overall, performance of SPD V1.0 in Wales follows the same broad patterns highlighted in the England and Wales comparisons presented above. In 2011, the total SPD V1.0 population estimate for Wales was 3,044,616 which is 0.61% lower than the census estimate for Wales in 2011. In 2013, the SPD V1.0 estimate of 3,069,146 was 0.43% lower than the census estimate.
MYE, and in 2014 the SPD V1.0 estimate of 3,072,726 was 0.62% lower than the MYE. Figure 4 below shows the five-year age and sex comparisons for Wales in 2011 compared to census estimates. There are some differences from the performance of the SPD for England and Wales as a whole, notably a more pronounced underestimate for children in the 0 to 4 age group, and a comparatively higher overestimate for males aged 35 to 39 (6% higher than the census estimate for Wales). Another difference from the England and Wales comparison in 2011 is an overestimate for males aged 25 to 29 on SPD V1.0.

**Figure 4, Difference between SPD V1.0 estimates and census estimates in 2011 by five-year age and sex group, Wales**

![Comparison of SPD v1.0 to 2011 Census by five-year age group and sex, Wales, 2011](image)

Source: Office for National Statistics

Figure 5 shows the same analysis for Wales in 2014 compared to the official MYE. The most notable change over the period between 2011 and 2014 is a more pronounced underestimate on the SPD for males and females aged 20 to 24, which was previously estimated reasonably close to the census in 2011.
An analysis for SPD V1.0 estimates for England has not been presented in this paper, as the comparisons are very similar to those observed for England and Wales as a whole. Overall, the results presented in this section demonstrate that there is considerable variation in the coverage of SPD V1.0 estimates at the national level when compared across five-year age and sex groups. While additional data sources that we are currently pursuing and their associated indicators of activity will help us develop new rules to improve on these coverage issues, the amount of variation apparent in this analysis would suggest that a Population Coverage Survey may be needed to help assess and adjust for coverage error in the SPD (see the Executive Summary of this document and our previous paper Estimating the Population: In Practice).
5.0 SPD V1.0 analysis at local authority level

This section looks at SPD V1.0 for 2011, 2013 and 2014 and compares these to the official estimates at local authority level. While at national level the 2011 Census had a confidence interval of ±0.15%, at the local authority level these confidence intervals averaged around 1.4% and ranged between 0.6% and 3.8%. For mid-year estimates at the mid-point of the last decade this interval is estimated to have been ±8.5%. These levels of accuracy have been used to derive the national P1 and P3 quality standards against which the SPDs have been assessed (see the Beyond 2011 Final Options report Paper O4 for more detail).

Table 7, Summary of quality standards for local authorities

<table>
<thead>
<tr>
<th>Quality standard</th>
<th>Population estimates (produced annually at the local authority (LA) level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 (maximum quality in the current system, ie. what is achieved in a census year)</td>
<td>All LA population estimates have a 95% confidence interval of ±3.8% or better</td>
</tr>
<tr>
<td>P3 (average quality in the current system)</td>
<td>All LA population estimates have a 95% confidence interval of ±8.5% or better (excluding outliers)</td>
</tr>
</tbody>
</table>

For individual age sex-groups, confidence intervals at the local authority level are wider still. Consequently, and with the inherent degradation in quality of the MYEs over time (see section 4.1), drawing conclusions about the quality of the SPD V1.0 based on such comparisons becomes more difficult. This further emphasises the need for an alternative method of assessing the quality of SPDs via a population coverage survey for example.

Note: These research outputs are NOT official statistics on the population
5.1 Difference between SPD V1.0 and official estimates in 2011

Figure 6 below shows the performance of SPD total population estimates at local authority level compared to census estimates. The current P1 quality standard is for 97% of local authority estimates to be within 3.8% of the census estimate.

Figure 6, Performance of 2011 SPD V1.0 total population estimates against P1 quality standard, local authorities
For SPD V1.0 2011, 94% (327) of the 348 local authorities in England and Wales are within the 3.8% (P1) quality standard. Of the 21 local authorities with SPD estimates outside of the quality standard, 18 were lower and 3 higher. Table 8 below shows the top 10 local authorities for SPD V1.0 total population estimates that are higher than census estimates in 2011. The three local authorities that estimated higher than the P1 quality standard are highlighted with an asterisk in this table.

### Table 8, LAs with higher SPD V1.0 estimates when compared to 2011 Census, All persons

<table>
<thead>
<tr>
<th>Rank</th>
<th>Local authority</th>
<th>Percentage difference</th>
<th>Total difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>City of London *</td>
<td>7.7%</td>
<td>570</td>
</tr>
<tr>
<td>2</td>
<td>Knowsley *</td>
<td>6.3%</td>
<td>9,180</td>
</tr>
<tr>
<td>3</td>
<td>Blackpool *</td>
<td>5.8%</td>
<td>8,280</td>
</tr>
<tr>
<td>4</td>
<td>Middlesbrough</td>
<td>3.6%</td>
<td>4,940</td>
</tr>
<tr>
<td>5</td>
<td>Torbay</td>
<td>3.6%</td>
<td>4,660</td>
</tr>
<tr>
<td>6</td>
<td>Plymouth</td>
<td>3.2%</td>
<td>8,210</td>
</tr>
<tr>
<td>7</td>
<td>Burnley</td>
<td>3.0%</td>
<td>2,630</td>
</tr>
<tr>
<td>8</td>
<td>Harlow</td>
<td>3.0%</td>
<td>2,440</td>
</tr>
<tr>
<td>9</td>
<td>Sunderland</td>
<td>2.8%</td>
<td>7,850</td>
</tr>
<tr>
<td>10</td>
<td>Lancaster</td>
<td>2.8%</td>
<td>3,920</td>
</tr>
</tbody>
</table>

* LAs with SPD estimates higher than the P1 quality standard (> 3.8% higher than 2011 Census estimate)

The three local authorities with differences higher than the 3.8% quality standard are City of London, Knowsley, and Blackpool. The higher estimates observed in City of London are possibly the result of high numbers of temporary residents in the area. City of London is unique in this respect as a local authority that is largely identifiable as a central business district. A potential explanation for the higher SPD V1.0 estimate is that the administrative data is capturing more inward migration as a result of these temporary residencies. Failures to update information when residents move out of the area has the effect of producing a higher population estimate. This is something that can be typically observed on the PR, where the number of patient registrations can be seen to rapidly accumulate each year, at a faster rate than they are removed when people leave the area. This process is discussed in more detail in the case study of Hammersmith and Fulham (see section 6.3).

To an extent, it is likely that both Blackpool and Knowsley suffer from the same issue relating to higher levels of inward migration being recorded on the administrative data. In Blackpool, SPD V1.0 is 5.8% (8,280 persons) higher than the census estimate, and the majority of this difference is for males (over 6,000 persons). Figure 7 below shows the sex-ratio in 2011 of males to females in Blackpool and Knowsley, for both the census and SPD V1.0. In Blackpool SPD V1.0, indicates an excess of working age males compared to females. As one of England’s larger seaside destinations, it might be reasonable to expect that administrative data sources accumulate records for people who are only temporarily resident due to seasonal working patterns. The higher SPD V1.0 estimate is likely to reflect the general tendency for males to take longer to register with GPs when leaving the area than females. As a result PR records continue to accumulate, producing higher SPD V1.0 estimates, particularly for males. This is discussed further in our case study of Hammersmith and Fulham in section 6.3.
In the case of Knowsley, the sex ratio of males to females is estimated by the 2011 Census to be one of the lowest in England and Wales, however this was also the case in 2001. While the sex ratio on SPD V1.0 appears plausible, the MYEs for 2013 have shown a net outward flow of young adults (particularly males) between the ages of 25 and 30. SPD V1.0
estimates continue to increase for this age, suggesting that list inflation may also be occurring in this local authority.

Table 9 below shows the top 10 local authorities for SPD V1.0 estimates that are lower than census estimates in 2011:

Table 9, LAs with lower SPD V1.0 estimates when compared to 2011 Census, All persons

<table>
<thead>
<tr>
<th>Rank</th>
<th>Local authority</th>
<th>Percentage difference</th>
<th>Total difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kensington and Chelsea</td>
<td>-21.9%</td>
<td>-34,780</td>
</tr>
<tr>
<td>2</td>
<td>Westminster</td>
<td>-16.7%</td>
<td>-36,530</td>
</tr>
<tr>
<td>3</td>
<td>Haringey</td>
<td>-9.4%</td>
<td>-24,050</td>
</tr>
<tr>
<td>4</td>
<td>Forest Heath</td>
<td>-8.6%</td>
<td>-5,110</td>
</tr>
<tr>
<td>5</td>
<td>Harrow</td>
<td>-7.2%</td>
<td>-17,150</td>
</tr>
<tr>
<td>6</td>
<td>Brent</td>
<td>-6.7%</td>
<td>-20,870</td>
</tr>
<tr>
<td>7</td>
<td>Camden</td>
<td>-6.3%</td>
<td>-13,940</td>
</tr>
<tr>
<td>8</td>
<td>Tunbridge Wells</td>
<td>-6.1%</td>
<td>-7,040</td>
</tr>
<tr>
<td>9</td>
<td>Barnet</td>
<td>-6.1%</td>
<td>-21,820</td>
</tr>
<tr>
<td>10</td>
<td>Hammersmith and Fulham</td>
<td>-5.9%</td>
<td>-10,860</td>
</tr>
</tbody>
</table>

All of the local authorities included in this table have SPD V1.0 estimates that fall well below the P1 quality standard. With the exception of Tunbridge Wells and Forest Heath, a number of London Boroughs are prominent in this table. These boroughs are characterised by high population churn and it is likely that delays in registering for services and notifying change of address have resulted in SPD V1.0 estimates that are lower than census estimates in 2011. For some local authorities the magnitude of difference between SPD V1.0 estimates and census estimates is considerable. Kensington and Chelsea, and Westminster have differences of -22% and -17% respectively.

Forest Heath is an example of a local authority with a high proportion of the population made up of armed forces personnel resident in military bases. While aggregate statistics are included in the SPD V1.0 estimate (see section 3.5) we currently have some uncertainty about the recording of dependents of armed forces personnel (for example children and spouses) on the PR. Consideration for this will be part of our ongoing work to better understand the availability of armed forces data in our future research.

The lower estimate produced by SPD V1.0 in Tunbridge Wells is more difficult to explain, as there are no apparent special population groups that can account for the difference from census estimates in 2011. Lower estimates on SPD V1.0 in Tunbridge Wells are prevalent across all five year age-sex groups in 2011, including males aged between 30 and 64.

The SPD V1.0 underestimate may partly be a consequence of net inward migration into the area, some of which has not been updated on the administrative records. As a consequence the administrative record is being counted elsewhere. In the case of Tunbridge Wells, this theory is partially supported by the fact that the proportion of half-weights (where records are weighted 0.5 at conflicting PR and CIS locations) are relatively high for a non-metropolitan local authority. It is also worth noting that the PR count for a number of age-sex groups in...
Tunbridge Wells are relatively low compared to census estimates. This may be an indication of particularly well maintained GP lists, or an area with an unusually high proportion of the population with alternative healthcare provision, meaning they are missed from the PR, and subsequently SPD V1.0.

5.2 Performance against P1 and P3 Quality Standards: 2011, 2013 and 2014

Table 10 below provides a summary of SPD comparisons with official estimates in 2011, 2013 and 2014.

Table 10, Performance of SPD V1.0 against P1 and P3 quality standards, LA Total Population Estimates

<table>
<thead>
<tr>
<th>Year</th>
<th>Performance within P1 Quality Standard (+/-3.8%)</th>
<th>SPD Performance within P3 Quality Standard (+/-8.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of local authorities</td>
<td>% of local authorities</td>
</tr>
<tr>
<td>2011</td>
<td>327</td>
<td>94.0%</td>
</tr>
<tr>
<td>2013</td>
<td>323</td>
<td>92.8%</td>
</tr>
<tr>
<td>2014</td>
<td>313</td>
<td>89.9%</td>
</tr>
</tbody>
</table>

The P1 quality standard requires that 97% of local authorities have a total population estimate within +/-3.8% of the 2011 Census estimate. This performance measure has also been used in 2013 and 2014, when comparing SPD V1.0 estimates to MYEs. In 2011, 94% (327) of local authorities had an SPD V1.0 estimate within 3.8% of the census estimate. The percentage of local authorities achieving this in subsequent years declined when compared to MYEs, with 92.8% in 2013 and 89.9% in 2014.

For the P3 quality standard, which requires that all local authorities have a total population estimate within +/-8.5% of the official estimate, performance is more consistent with 99% of local authorities within the quality standard in 2011, 2013 and 2014. Those local authorities with SPD V1.0 estimates outside of the P1 and P3 quality standards are largely made up of London Boroughs.

Table 11 below provides a breakdown of performance against the quality standards for England and Wales separately. Of particular note is the strong and consistent performance of SPD V1.0 in Wales where all 22 local authorities achieve the P1 and P3 quality standard in 2011 and 2013. Gwynedd is the only local authority in Wales to fall outside the P1 quality standard in 2014. In this case the SPD V1.0 estimate was -4.28% lower than the MYE in 2014, and this difference is largely the result of large population increases picked up in the MYE over the period.

Note: These research outputs are NOT official statistics on the population
Table 11, Performance of SPD V1.0 against P1 and P3 quality standards, local authority total population estimates, England and Wales

<table>
<thead>
<tr>
<th>Year</th>
<th>Welsh local authorities</th>
<th>English local authorities</th>
<th>Welsh local authorities</th>
<th>English local authorities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of LAs</td>
<td>% of LAs</td>
<td>No. of LAs</td>
<td>% of LAs</td>
</tr>
<tr>
<td>2011</td>
<td>22</td>
<td>100%</td>
<td>305</td>
<td>93.6%</td>
</tr>
<tr>
<td>2013</td>
<td>22</td>
<td>100%</td>
<td>301</td>
<td>92.3%</td>
</tr>
<tr>
<td>2014</td>
<td>21</td>
<td>95.5%</td>
<td>292</td>
<td>89.6%</td>
</tr>
</tbody>
</table>

5.3 SPD V1.0 in London local authorities

The age structure of the population in London, with a distinct peak for those aged around 30, indicates that a large number of people arrive in their 20s and a similarly large number of people leave in their 30s and 40s. Local authorities with high levels of net migration are more prone to lags in the registration of inflows of the population, and this often results in an underestimate of those in their 20s. Similarly, lags on capturing the outflows leads to an overestimate of those in their 30s and 40s. This is a particular problem in inner London and leads to SPD V1.0 estimates which typically substantially underestimate those in their 20s, broadly estimate well for ages 30-39 and then slightly overestimate those in their 40s and 50s.

Figure 8, SPD V1.0 and 2011 Census estimates, London, Males

Note: These research outputs are NOT official statistics on the population.
Figure 8 above shows how the SPD estimate in London is initially low for age groups under 35, but is subsequently higher for age groups between 40 and 59. In circumstances where a local authority experiences strong population growth it is likely that the SPD will underestimate the size of the population. This is largely due to delays in individuals registering with GPs or updating changes of address on the CIS (via DWP or HMRC), and is most prevalent amongst younger age groups. In these circumstances, out of date address information on the PR and CIS has the effect of underestimating population inflows into these local authorities, which will at best be delayed until an interaction with services is recorded, or in some cases wholly missed. Thus for London, those in their 20s are systematically underestimated as there tends to be a higher population inflow than outflow for this age group. By around age 30 however, net migration for inflows and outflows tends to be more even in London resulting in SPD estimates that approximate the population more accurately. For age groups over 40, net migration tends to be characterised by higher outflows from London, and these conditions lead to overestimation in the SPD estimates.

To some extent the dynamics that affect the SPDs also affect the official population estimates, as a large proportion of change in the official estimates is driven by change in the patient register (through internal migration and through the apportionment of international immigration to local authorities using administrative data). The challenges of estimating migration in the MYE methodology is summarised in ONS user update papers for internal migration and international migration. A key difference is that the official estimates are periodically recalibrated using a census meaning that they can ‘drift’ for 10 years at most. By comparison the SPD series in some local authorities has the potential to be in a state of permanent ‘drift’.

There are other factors that are likely to be impacting SPD performance in areas of low international and internal migration. Further research is needed to ascertain how special population groups, for example prisoners and school boarders, are recorded on the PR. Institutions housing special population groups will often be providing independent health services, which are not recorded on the PR. These conditions have the potential to underestimate for particular age groups in areas where these special populations are most prevalent.

5.4 SPD V1.0 Performance by five year age-sex at local authority level

Overall, the method used in the construction of SPD V1.0 produces estimates for the majority of local authorities that are similar in level and age-sex distribution to those produced by the official estimates. SPD V1.0 estimates for females are generally of a very high quality (i.e. they are very closely aligned to official estimates) while estimates for males tend to be less well aligned. The remainder of this chapter will now focus on the performance of SPD V1.0 at the five year age-sex level. We have targeted our analysis towards areas where differences are greatest between SPD V1.0 estimates and official estimates, to offer some potential explanations as to why some of these differences occur. We are particularly interested in receiving feedback from users of these research outputs that have further insight into the possible causes of these discrepancies or thoughts about how
the method could be developed to improve the SPD estimates. Users are encouraged to do so via the template available here, or by contacting Beyond.2021.Research.and.Design@ons.gov.uk

While SPD V1.0 performed reasonably well against the quality standards for LA total populations, standards have also been set to evaluate performance of LA estimates at five year age-sex level. Focussing on total population estimates has the potential to hide some of the coverage issues that occur on SPD V1.0 for particular age-sex groups. Section 4 has already shown that at the national level there is considerable variability across five year age-sex groups, with higher estimates produced for working age males and notably lower estimates for other age groups. This variability will also be evident at local authority level, and in some instances the cancelling effect of overestimation and underestimation across age-sex groups will result in a good approximation for the total population estimate.

Table 12 below shows the number of local authorities that are within the P1 and P3 quality standards for each five year age-sex group in 2011.

Table 12, SPD V1.0 performance against P1 and P3 quality standards in 2011

<table>
<thead>
<tr>
<th>Five Year Age Group</th>
<th>% LAs within P1 Quality Standard (+/-7.2%)</th>
<th>% LAs within P3 Quality Standard (+/-12.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td>0 to 4</td>
<td>95.4%</td>
<td>94.8%</td>
</tr>
<tr>
<td>5 to 9</td>
<td>94.0%</td>
<td>92.8%</td>
</tr>
<tr>
<td>10 to 14</td>
<td>89.1%</td>
<td>89.9%</td>
</tr>
<tr>
<td>15 to 19</td>
<td>87.6%</td>
<td>86.8%</td>
</tr>
<tr>
<td>20 to 24</td>
<td>72.7%</td>
<td>83.1%</td>
</tr>
<tr>
<td>25 to 29</td>
<td>64.4%</td>
<td>86.2%</td>
</tr>
<tr>
<td>30 to 34</td>
<td>75.6%</td>
<td>94.8%</td>
</tr>
<tr>
<td>35 to 39</td>
<td>74.4%</td>
<td>96.0%</td>
</tr>
<tr>
<td>40 to 44</td>
<td>81.0%</td>
<td>96.8%</td>
</tr>
<tr>
<td>45 to 49</td>
<td>86.8%</td>
<td>98.3%</td>
</tr>
<tr>
<td>50 to 54</td>
<td>92.0%</td>
<td>97.7%</td>
</tr>
<tr>
<td>55 to 59</td>
<td>96.6%</td>
<td>98.6%</td>
</tr>
<tr>
<td>60 to 64</td>
<td>98.6%</td>
<td>98.3%</td>
</tr>
<tr>
<td>65 to 69</td>
<td>98.9%</td>
<td>98.9%</td>
</tr>
<tr>
<td>70 to 74</td>
<td>98.6%</td>
<td>98.6%</td>
</tr>
<tr>
<td>75 to 79</td>
<td>99.1%</td>
<td>98.6%</td>
</tr>
<tr>
<td>80 to 84</td>
<td>98.0%</td>
<td>97.7%</td>
</tr>
<tr>
<td>85+</td>
<td>95.4%</td>
<td>96.8%</td>
</tr>
</tbody>
</table>

The P1 quality standard for five-year age groups is for at least 97% of local authorities to be within 7.2% of the census estimate. Table 7 shows that SPD V1.0 is close to achieving that for some age groups, but well below the standard in younger adult age groups, particularly for males.

Note: These research outputs are NOT official statistics on the population
An interactive map with accompanying population pyramids for SPD V1.0 estimates has been provided to accompany these research outputs. These are available [HERE](#) to provide more detailed analysis of SPD V1.0 performance at local authority level.

In most cases, local authorities that have large differences for five year age groups when comparing SPD V1.0 estimates with official estimates will also have a large difference in the comparison of total populations. However there are some local authorities with small net differences between population totals and large differences for each age group. In assessing the similarity between SPDs and official estimates we have used a measure of average absolute difference.

*Average absolute difference:* This approach assesses how similar population estimates are from two different data sources. Figure 9 below is based on a calculation of average absolute differences from 2011 Census estimates, for both SPD estimates and the rolled forward MYE in 2011. This measure is calculated for each five-year age sex-group as the average absolute difference from the census estimate from all local authorities. The chart shows how SPD V1.0 would have outperformed the MYE method for most age groups should it have been carried forward without a census rebase in 2011\(^7\). The peak difference for the rolled forward MYEs is observed amongst 20 to 24 year old males where on average local authorities differed from the census estimate by about 500 persons. This compares with an average of 350 persons on SPD V1.0. The peak difference on SPD V1.0 is for males aged 25 to 29 where on average local authorities differ by about 400 persons from the census estimate. Differences in the rolled forward MYE are still higher for this age group, at about 500 persons.

---

\(^7\) Note that the rolled forward 2011 MYEs were produced for reference date 30\(^{th}\) June, which will account for some of the differences from the 2011 Census day 27\(^{th}\) March

Note: These research outputs are NOT official statistics on the population
As explained in section 4.2, accuracy of the MYEs tends to drift towards the end of the intercensal period. Figure 6 shows that SPD V1.0 estimates in 2011 are generally closer to census estimates when compared to a mid-year estimate that has been rolled forward into its 10th year. It should be noted however that for the majority of the intercensal period the MYEs would outperform SPD V1.0 if comparisons were available.

Table 13 below shows that in some instances there can be small differences in the total SPD V1.0 population estimate when compared to 2011 Census, but larger differences when averaged over five-year age groups. In this instance, the average absolute difference is calculated for each local authority separately and across all five-year age-sex groups.
### Table 13, Local authorities with highest average absolute differences across five year age group, SPD and 2011 Census

<table>
<thead>
<tr>
<th>Rank</th>
<th>Local authority</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average absolute difference across 5 year age groups (%)</td>
<td>Absolute difference in total population (%)</td>
</tr>
<tr>
<td>1</td>
<td>City of London</td>
<td>22%</td>
<td>10%</td>
</tr>
<tr>
<td>2</td>
<td>Kensington and Chelsea</td>
<td>21%</td>
<td>21%</td>
</tr>
<tr>
<td>3</td>
<td>Westminster</td>
<td>17%</td>
<td>16%</td>
</tr>
<tr>
<td>4</td>
<td>Lambeth</td>
<td>12%</td>
<td>1%</td>
</tr>
<tr>
<td>5</td>
<td>Hackney</td>
<td>12%</td>
<td>4%</td>
</tr>
<tr>
<td>6</td>
<td>Haringey</td>
<td>12%</td>
<td>11%</td>
</tr>
<tr>
<td>7</td>
<td>Isles of Scilly</td>
<td>12%</td>
<td>1%</td>
</tr>
<tr>
<td>8</td>
<td>Forest Heath</td>
<td>12%</td>
<td>4%</td>
</tr>
<tr>
<td>9</td>
<td>Richmondshire</td>
<td>11%</td>
<td>4%</td>
</tr>
<tr>
<td>10</td>
<td>Wandsworth</td>
<td>10%</td>
<td>3%</td>
</tr>
<tr>
<td>11</td>
<td>Knowsley</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>12</td>
<td>Hammersmith and Fulham</td>
<td>10%</td>
<td>8%</td>
</tr>
<tr>
<td>13</td>
<td>Camden</td>
<td>9%</td>
<td>7%</td>
</tr>
<tr>
<td>14</td>
<td>Blackpool</td>
<td>9%</td>
<td>9%</td>
</tr>
<tr>
<td>15</td>
<td>Islington</td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>16</td>
<td>Southwark</td>
<td>8%</td>
<td>1%</td>
</tr>
<tr>
<td>17</td>
<td>Brent</td>
<td>8%</td>
<td>7%</td>
</tr>
<tr>
<td>18</td>
<td>Oxford</td>
<td>8%</td>
<td>5%</td>
</tr>
<tr>
<td>19</td>
<td>Bristol, City of</td>
<td>8%</td>
<td>2%</td>
</tr>
<tr>
<td>20</td>
<td>Manchester</td>
<td>8%</td>
<td>1%</td>
</tr>
</tbody>
</table>

There are a small number of local authorities where SPD V1.0 total populations are similar but with notably larger differences when they are averaged over five-year age groups. For example, the total population difference between SPD V1.0 and 2011 Census estimate for males in Lambeth was only 1%, whereas the average absolute difference for males across 5 year groups in Lambeth was 12%. The SPD V1.0 total for males in Lambeth is therefore the result of large discrepancies across the five-year age groups (high and low) that have cancelled out to produce a population total that is relatively close to the 2011 Census estimate. In the majority of cases however, local authorities with the greatest differences in total populations also have the highest average absolute differences for five-year age groups.

Several other London local authorities, Hackney, Southwark and Wandsworth, have total population estimates on SPD V1.0 that are similar to census despite there being substantial differences in the age-sex distributions on SPD V1.0. These local authorities all have high levels of population migration. The process that leads to population estimates for some age groups being underestimated and others being overestimated in London authorities is

Note: These research outputs are NOT official statistics on the population.
described in section 5.3. This is primarily a problem with estimates for males, reflecting the generally higher quality of administrative data for females, which tends to result in smaller differences observed between SPD V1.0 estimates and official estimates. A comparison between the two charts in figures 10 and 11 below shows the extent to which larger differences across five-year age groups are observed for males than females. The charts also demonstrate that for males, these large differences across five-year age groups are more likely to result in small differences in total population comparisons.

Figure 10, SPD V1.0 total population difference compared to absolute average difference for five-year age groups, males, 2011

Source: Office for National Statistics
5.5 Differences between SPD V1.0 and official estimates in 2011 and 2014

In this section we consider whether the differences observed between SPD V1.0 estimates and official estimates are stable over time. For a small number of local authorities, the differences between total SPD V1.0 population estimates and official population estimates change considerably between 2011 and 2014. Table 14 below shows the top 10 local authorities with the highest SPD V1.0 estimates compared to the MYE in 2014. It also shows the difference observed for these local authorities when the 2011 SPD V1.0 estimates are compared with census estimates in 2011.
Administrative Data Research Report: 2015

Table 14, Top 10 LAs where SPD V1.0 estimates are higher than MYEs in 2014, All persons

<table>
<thead>
<tr>
<th>Rank</th>
<th>Local authority</th>
<th>% higher on 2014 SPD V1.0 compared to 2014 MYE</th>
<th>Estimate higher on 2014 SPD V1.0 compared to 2014 MYE</th>
<th>% higher on 2011 SPD V1.0 compared to 2011 Census</th>
<th>Estimate higher on 2011 SPD V1.0 compared to 2011 Census</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Blackpool</td>
<td>6.2</td>
<td>8,700</td>
<td>5.8</td>
<td>8,300</td>
</tr>
<tr>
<td>2</td>
<td>Knowsley</td>
<td>6.0</td>
<td>8,800</td>
<td>6.3</td>
<td>9,200</td>
</tr>
<tr>
<td>3</td>
<td>Peterborough</td>
<td>5.1</td>
<td>9,700</td>
<td>1.3</td>
<td>2,500</td>
</tr>
<tr>
<td>4</td>
<td>Middlesbrough</td>
<td>4.2</td>
<td>5,800</td>
<td>3.6</td>
<td>4,900</td>
</tr>
<tr>
<td>5</td>
<td>Boston</td>
<td>3.9</td>
<td>2,600</td>
<td>1.9</td>
<td>1,200</td>
</tr>
<tr>
<td>6</td>
<td>Wellingborough</td>
<td>3.8</td>
<td>2,900</td>
<td>2.1</td>
<td>1,600</td>
</tr>
<tr>
<td>7</td>
<td>Reading</td>
<td>3.6</td>
<td>5,900</td>
<td>1.2</td>
<td>1,800</td>
</tr>
<tr>
<td>8</td>
<td>Ealing</td>
<td>3.3</td>
<td>11,400</td>
<td>-2.1</td>
<td>-7,000</td>
</tr>
<tr>
<td>9</td>
<td>Blackburn with Darwen</td>
<td>3.3</td>
<td>4,900</td>
<td>1.4</td>
<td>2,100</td>
</tr>
<tr>
<td>10</td>
<td>Burnley</td>
<td>3.3</td>
<td>2,900</td>
<td>3.0</td>
<td>2,600</td>
</tr>
</tbody>
</table>

Generally, most of these local authorities had higher estimates when SPD V1.0 was compared in 2011 with census estimates, however in some cases the difference has notably increased. Ealing is an example of where in 2011 the SPD V1.0 estimate was actually lower than the census estimate by 2.06%, whereas in 2014 it was higher than the MYE by 3.3%.

Table 15 below shows the top 10 local authorities with the lowest SPD V1.0 estimates compared to the MYE in 2014. With the exception of Forest Heath, and Kensington and Chelsea, the difference from the official estimates was larger in 2014 than in 2011. In some cases the difference has notably increased, for example in Exeter, the Isles of Scilly, and the London Boroughs of Camden, Southwark and Tower Hamlets.

Table 15, Top 10 LAs where SPD V1.0 estimates are lower than MYEs in 2014, all persons

<table>
<thead>
<tr>
<th>Rank</th>
<th>Local authority</th>
<th>% lower on 2014 SPD V1.0 compared to 2014 MYE</th>
<th>Estimate lower on 2014 SPD V1.0 compared to 2014 MYE</th>
<th>% lower on 2011 SPD V1.0 compared to 2011 Census</th>
<th>Estimate lower on 2011 SPD V1.0 compared to 2011 Census</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Westminster</td>
<td>-19.5</td>
<td>-45,500</td>
<td>-16.7</td>
<td>-36,500</td>
</tr>
<tr>
<td>2</td>
<td>Kensington and Chelsea</td>
<td>-18.7</td>
<td>-29,200</td>
<td>-21.9</td>
<td>-34,800</td>
</tr>
<tr>
<td>3</td>
<td>Isles of Scilly</td>
<td>-13.6</td>
<td>-300</td>
<td>-2.3</td>
<td>-100</td>
</tr>
<tr>
<td>4</td>
<td>Camden</td>
<td>-11.3</td>
<td>-26,400</td>
<td>-6.3</td>
<td>-13,900</td>
</tr>
<tr>
<td>5</td>
<td>Southwark</td>
<td>-7.9</td>
<td>-23,900</td>
<td>-3.1</td>
<td>-9,000</td>
</tr>
<tr>
<td>6</td>
<td>Forest Heath</td>
<td>-7.8</td>
<td>-4,900</td>
<td>-8.6</td>
<td>-5,100</td>
</tr>
<tr>
<td>7</td>
<td>Tower Hamlets</td>
<td>-7.4</td>
<td>-21,100</td>
<td>-3.9</td>
<td>-9,900</td>
</tr>
<tr>
<td>8</td>
<td>Islington</td>
<td>-6.7</td>
<td>-14,700</td>
<td>-4.3</td>
<td>-8,800</td>
</tr>
<tr>
<td>9</td>
<td>Exeter</td>
<td>-6.6</td>
<td>-8,200</td>
<td>0.3</td>
<td>400</td>
</tr>
<tr>
<td>10</td>
<td>Rutland</td>
<td>-6.0</td>
<td>-2,300</td>
<td>-5.4</td>
<td>-2,000</td>
</tr>
</tbody>
</table>

Changes observed in the relative differences to official estimates over time are difficult to account for with certainty. Much of this change can be attributed to changes in the administrative data, however, in some instances there are also issues relating to the method.

Note: These research outputs are NOT official statistics on the population
for producing MYEs. It is crudely possible to categorise local authorities included into groups experiencing similar issues. We have identified five groups of LAs within tables 11 and 12 as the focus of further discussion in this section of the report. Generally the similarity between estimates for males changes the most between 2011 and 2014 in comparison to females. Figure 12 focuses on an analysis of males, however many of the same patterns are observable in females for the same local authorities.

**Figure 12, Percentage difference between SPDs and official estimates in 2011 and 2014, males**

The first group identified in the analysis comprises west London boroughs (Brent, Harrow and Ealing). By observing differences in PR extracts in years 2011, 2012 and 2013, it appears that these areas may have undergone a GP list clean in 2011, which was evidenced by a higher proportion of ‘FP69’ flags being assigned to patient records that year. An FP69 flag is an indication on the patient register that the person has been sent correspondence to confirm whether they are still living at the address recorded. This process is the subject of more detailed analysis in our case study of Charnwood in section 6.5.
Figure 13 below shows the extent to which FP69 flags were evident on the PR for Brent in 2011, compared to 2013 and 2014.

**Figure 13, FP69 allocation on the Patient Register, 2011, 2013 and 2014, Brent, males**

It is clear that in 2011 a much higher number of FP69s were recorded on the PR, reflecting an increased effort at the time to contact patients to confirm whether they were still resident in the area. Non-response to this correspondence usually results in removal from the GP list, so we have used the FP69 flag as one of our rules to exclude individuals from the SPD. The exclusion of these records from SPD V1.0 may offer some explanation for the lower estimates in Brent (as well as neighbouring authorities Harrow and Ealing) relative to the official estimates. In 2013 and 2014 the effect of this list cleaning exercise has diminished and the SPD V1.0 estimates for these LAs have 'bounced back'. In 2011 SPD V1.0 estimates for these LAs were among the most underestimated in England and Wales, by 2013 they were overestimating.

The second group comprises Devon and Cornwall. In 2011, Devon and Cornwall had SPDs that were broadly equivalent to their official estimates, with the notable exception of Torbay which was around 3.6% higher than the census estimate. In 2014, Devon and Cornwall were undergoing a patient register list cleaning exercise which generated a large number of FP69s, which resulted in these records being removed from the SPD V1.0 (see published Pulse article). Consequently the SPD V1.0 estimates for 2014 are systematically underestimated relative to the official estimates. We are reviewing the use of FP69 flags as the basis for removing records from the SPD in future years. It may be the case that the...
The third group is comprised of Boston, Wellingborough and Peterborough. These are local authorities where the increase in the SPD has been faster than the increase in the official estimates. In each case the SPD was already higher than the official estimates in 2011 (by 1.9% for Boston, 2.1% for Wellingborough and 1.3% for Peterborough) and this difference has increased (to 3.9%, 3.8% and 55.1% respectively). For each of these local authorities part of the explanation is probably due to international migration. This could be either the result of the SPD including a large proportion of short-term international economic migrants or the official estimate of net international migration for these areas being underestimated. For Wellingborough part of the discrepancy is possibly related to the closure of HMP Wellingborough in 2012. Prisoners in HMP Wellingborough would have been included in the MYE population but we have uncertainty regarding the updating of prisoners’ details on the PR and CIS following this closure. This may explain why estimates on SPD V1.0 are higher in 2013 and 2014 however we need to undertake further research to understand the processes for registering and updating administrative records for prisoners.

The fourth group is local authorities with large numbers of students. The mid-year estimates measure of internal migration relies on moving graduates out of their local authority of study by detecting moves between extracts on the PR. Lags in the process for graduates registering with GPs when leaving the area of study has the potential to have overestimated the number of 20 to 24 year olds in student LAs in 2013 and 2014 MYEs. In 2011, SPD V1.0 generally underestimated the population age 20 to 24 in student local authorities, and the level of underestimation may have appeared to have increased due to inflated estimates for this age group in the MYE.

The fifth group to consider is those local authorities with large proportions of armed forces. Forest Heath has already been identified above as having a notably lower estimate on SPD V1.0 when compared to census estimates in 2011. However, between 2011 and 2014 the differences between SPD V1.0 and official population estimates have increased for other local authorities with a high proportion of armed forces in the population. For example Richmondshire’s 2011 SPD V1.0 estimate was just 1.5% lower than the 2011 Census estimate, while in 2014 this had increased to 4.6% when compared to the 2014 MYE. At least some of this increase is due to using official estimates that are based on different residential definitions. The 2011 Census put armed forces personnel at their home/family address where possible. Conversely the mid-year estimates put armed forces personnel in the LA where they are based. This means that the SPD V1.0 estimates appear more underestimated relative to mid-year estimates than they do when compared to a census estimate.

There are two outlying local authorities in figure 10 (Westminster, and Kensington and Chelsea) where SPD V1.0 estimates were below the census estimate in 2011. In the case of Westminster, the difference has increased from -16.7% when compared to the census estimate in 2011, to -19.5% when compared to the MYE for 2014. In Kensington and Chelsea, the gap between official estimates has decreased, from -21.9% when compared to the census estimate in 2011, to -18.7% when compared to the MYE in 2014.
5.6 Methodological considerations for changes in SPD V1.0 and MYEs

The main distinguishing features between the two methods have been broadly outlined in section 3. To summarise, official population estimates are based on components of change whereas SPDs reflect total change over time.

In comparing change over time from the SPDs and the official series it is important to fully understand the differences between the two series. In broad terms the change over time seen in the official series is mirrored by similar change in the SPDs suggesting that they are generally measuring change in the same phenomena. There are however some differences. As discussed earlier the SPD series is occasionally subject to sudden changes in level as a consequence of PR list cleaning, this means that change in the SPD series is a combination of real population change and artefactual change due to administrative processes. Two examples of this issue, Cornwall and Brent are shown in figure 10 below. In the case of Cornwall, overall change in the number of patient register records was relatively flat between 2011 and 2014. However, our reliance on FP69 flags as the basis for removal (described in section 4.3) appears to have resulted in undue exclusions from the SPD in 2014. In reality the level of change on the patient register appears to mirror the real population change measured in the MYEs.

In the case of Brent, artefactual change in the administration of FP69 flags has had the opposite effect. Following an attempt to reduce list inflation on the patient register through the FP69 process, it is believed that a proportion of those targeted were retained on the list having confirmed that they were still resident. As a result FP69s were removed in subsequent extracts and the change in SPD estimates is far higher than true population change as a consequence.
Figure 14, Examples of artefactual change in SPDs, Cornwall and Brent

**Population change between 2011 and 2014 by five-year age group**

**Cornwall**

**Males**

![Graph](image)

*Source: Office for National Statistics*

**Population change between 2011 and 2014 by five-year age group**

**Brent**

**Males**

![Graph](image)

*Source: Office for National Statistics*

Note: SPD V1.0 data and Patient register data for 2011 refer to census day 2011 (27 March) while SPD V1.0 and PR data for 2013 and 2014 refer to mid-year (30 June)

Note: These research outputs are NOT official statistics on the population
Change over time measured by the official estimates and the SPDs for student areas exhibit a number of differences reflecting issues with the accurate measurement of graduates leaving university areas in the official estimates. Since 2011, challenges in the measurement of internal migration used in the mid-year estimates, particularly for moving out graduates from local authorities with large student populations, have been identified. This has resulted in official estimates of the number of 20-24 year olds in local authorities with large student populations increasing. Improvements that have been made in the estimation of student migration in the MYEs is summarised in this Improving Migration and Population Statistics paper. SPD V1.0 estimates for the movement of students differently, students are located at their HESA term-time address and then revert to their PR or CIS address(es) when they are no-longer on HESA. In some instances, this data will be out of date, reflecting the address recorded before studying at university. Looking at evidence from 2011 this may lead to a tendency for the SPDs to underestimate those aged 20-24 in local authorities with large student populations. This can be observed in each of the charts provided in figure 15 below.

Figure 15, Local authorities with students, population change from different sources

Source: Office for National Statistics

Note: These research outputs are NOT official statistics on the population
Population change between 2011 and 2014 by five-year age group
Nottingham
Males

Source: Office for National Statistics

Population change between 2011 and 2014 by five-year age group
Oxford
Males

Source: Office for National Statistics

Note: These research outputs are NOT official statistics on the population
5.7 SPD V1.0: Summary of overall performance

The method used to produce SPD V1.0 for 2011 produces estimates for the majority of local authorities that are similar in level and age/sex distribution to those produced by the 2011 Census. Furthermore it should be noted that SPD V1.0 estimates for females are generally of a very high quality (very closely aligned to official estimates) while estimates for males tend to be less well aligned. While we have highlighted that SPD V1.0 underestimates at the national level for those aged 85+, generally it produces estimates for older age groups that are very similar to the official estimates.

In some local authorities the SPD V1.0 estimates are significantly different to the official estimates. Often these differences may be explained by the presence in the official estimates of population sub-groups not covered adequately by the SPD V1.0 approach, most notably armed forces personnel, prisoners and boarding school pupils. More research is needed to understand how these population sub-groups interact with the administrative data used in the construction of SPD V1.0. Where local authorities have large numbers of these types of people, the SPD V1.0 estimate is likely to notably underestimate the population. In local authorities with positive net internal migration for those in age 20-29 and negative net internal migration for those in their 30s and 40s, a pattern often seen in inner London, the SPD V1.0 method produces estimates that are significantly different to those from the census and mid-year estimates. This reflects the tendency for delays in the registration and updating of addresses on both the patient register and the CIS, as individuals are slow to interact with administrative systems following a move.
A consequence of the particular pattern and high levels of migration in inner London, related to the life cycle, is the greater relative impact of delays in updating administrative data used in the construction of SPD V1.0. SPD V1.0 estimates are systematically lower than official estimates across inner London for those aged 20-29 reflecting the high levels of inward migration in this age group.

Between 2011 and 2014, the differences between SPD V1.0 estimates and official estimates have widened. This pattern of change is to be expected and partly reflects the reduction in the quality of the official estimates between censuses. In addition the relationship is affected by the periodic episodes of PR list cleaning that occur in different parts of the country, which can result in large reductions from the SPD V1.0 estimate, for example, Brent in 2011 and Cornwall in 2014. In addition the different residential definitions used in the mid-year estimates and the 2011 Census, particularly in relation to special populations such as prisoners and armed forces personnel, can result in SPD V1.0 estimates that are further from official estimates in 2013 and 2014 than in 2011. This is particularly evident in local authorities with large armed forces populations but can also be seen in areas with relatively high populations of school boarders or prisoners.

The next section of this report focuses on the analysis of quality relating to the administrative datasets used to compile SPD V1.0. A series of case studies have been undertaken to highlight particular issues that are expected to have caused discrepancies between the SPD estimates and the official estimates.
**6.0 Quality of administrative data**

This section of the paper includes a more detailed analysis of selected local authorities to highlight underlying issues associated with the statistical quality of the administrative data. While government departments have procedures in place to ensure that data is of high quality to manage services provided, there are wider quality issues associated with the use of these data in the production of population statistics. Quality issues affecting the relationship between SPD estimates and MYEs relate to a number of areas. These are summarised in the table below.

Table 16, Impact of different issues with SPDs, census and mid-year estimates on the relationship between SPDs and official population estimates

<table>
<thead>
<tr>
<th>Issue</th>
<th>Type</th>
<th>Affect on relationship between official estimates and SPD V1.0 at local authority level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census confidence intervals</td>
<td>Statistical uncertainty</td>
<td>SPD V1.0 and official estimates will differ due to sampling error around census estimates, this will affect some age/sex groups and local authorities more than others reflecting differences in the precision of census estimates.</td>
</tr>
<tr>
<td>National discrepancy between SPD and official estimates</td>
<td>Coverage</td>
<td>SPD V1.0 is generally lower than official estimates. SPD V1.0 for some age/sex groups is higher than the official estimates (typically males aged between 30 and 64).</td>
</tr>
<tr>
<td>Armed forces</td>
<td>Different geographic distribution in official population estimates</td>
<td>The treatment of armed forces personnel on the census differs from their treatment in the mid-year estimates; the census enumerated armed forces personnel at their family home where possible where as the MYEs account for them at their home military base. The SPD V1.0 series accounts for armed forces in the same way to the MYEs, consequently SPD V1.0 for armed forces areas in 2011 can appear overestimated.</td>
</tr>
<tr>
<td>School boarders</td>
<td>Coverage</td>
<td>SPD V1.0 estimates for areas with large numbers of boarding school pupils will be underestimated as these are unlikely to be fully represented on the PR or CIS in these areas.</td>
</tr>
<tr>
<td>Prisoners</td>
<td>Coverage</td>
<td>SPD V1.0 estimates for areas with large numbers of prisoners will be underestimated as these are unlikely to be fully represented on the PR or CIS in these areas.</td>
</tr>
<tr>
<td>Census</td>
<td>Age distribution</td>
<td>The 2011 Census over imputed the ages of some children at the expense of other ages. This affects all local authorities in England and Wales and means that SPD V1.0 for some ages are often overestimated or underestimated relative to official estimates.</td>
</tr>
<tr>
<td>Imputation</td>
<td>Geographic distribution/coverage</td>
<td>The SPD V1.0 method produces estimates for those of student age in student local authorities that are lower than the equivalents for the official estimates. We are exploring whether the SPD V1.0 rules are currently excluding a number of foreign students that should be included in the population estimate.</td>
</tr>
</tbody>
</table>

Note: These research outputs are NOT official statistics on the population
Graduates in 2014          Geographic distribution: The gap between official estimates and SPD V1.0 for those aged 20 to 24 in student local authorities has increased between 2011 and 2014. An issue with the measurement of internal migration of graduates has caused the estimates for those aged 20 to 24 in the mid-year estimates to become overestimated.

Lagging          Geographic and age distribution: Where populations are subject to large scale population change administrative data often takes time to catch up, this means that in areas/age groups subject to large population increases or decreases, SPD V1.0 tends to understate the amount of change. This is particularly the case for young adult males in urban areas but can also affect other groups.

Patient register list cleaning          Level: Periodic list cleaning of the patient register can cause SPD V1.0 for some local authority to decrease in size between years. In the years following a list cleaning exercise SPD V1.0 can recover as registrations resume accumulating.

To draw out some of the specific issues relating to the construction of SPD V1.0 and the administrative data underpinning it, we present four case studies to alert users to some of the typical problems observable amongst local authorities when comparing to official estimates. We would be interested to hear from users, particularly relating to the areas highlighted here, who can provide additional insight into our findings. Users are encouraged to do so via the template available HERE, or by contacting Beyond.2021.Research.and.Design@ons.gov.uk

The first is an analysis of the impact of using a half-weight approach for distributing SPD records across local authorities. The use of half weights is somewhat arbitrary but in the absence of other information it is an arguably reasonable approach to estimating people who have a different address recorded on the PR to that on the CIS. This difference is likely to be caused by a lag that has occurred in updating addresses on either of these sources. The implications of using this method is the potential for error in distributing records on the SPD, particularly if some local authorities have a higher propensity for address update lags than others. We explore this in some detail in the case of South Bucks.

The second case study takes a more detailed look at an example of list inflation that appears to be occurring on the PR. This demonstrates the vulnerability of an SPD method that relies on stock estimates from an administrative source that is rapidly accumulating large numbers of records as a result of high migration churn and individuals not interacting with service providers to notify when they have moved from the area. We have undertaken an analysis of Hammersmith and Fulham to explore this issue further.

The third case study explores the differences between SPDs and MYEs in an area populated with university students. A number of local authorities with large university halls and student accommodation have a tendency for lower estimates on the SPD for persons aged 20-24. Since the methodology for SPD V1.0 relies on assigning students wholly to HESA term time locations, the tendency for lower estimates cannot be attributed to half weights. We have undertaken an analysis in Cambridge to investigate possible reasons why SPD estimates for these age groups are notably lower than the official census estimates in 2011.

Note: These research outputs are NOT official statistics on the population
The fourth case study provides a more detailed example of the effects of patient register list cleaning. Sections 4 and 5 made reference to our use of FP69 flags to remove records from the SPD on the basis that they are unlikely to be resident at the address recorded. This flag is typically assigned to individuals who have been out of contact with their GP for around a year, although in practice the process for doing this varies considerably between areas. There is evidence that this list cleaning process in some areas can be both infrequent and intensive when undertaken. As a consequence, considerable differences between SPD V1.0 and official estimates are observed as the patient register drops significantly one year, before steadily re-inflating in the subsequent years. In this context, we present a case study based on Charnwood.

6.1 Quality indicators from administrative data

A number of quality indicators have been derived from the administrative sources to aid the interpretation of differences between SPD estimates and MYEs. These indicators relate to how the administrative data has been used in constructing the SPD, and can be indicative of issues relating to coverage and accuracy. These measures do not imply anything about the suitability of the administrative data for the purpose which they are collected, but about their quality in the context of population statistics.

The following indicators have been used as the basis for analysis:

- Estimate of people flagged as out of contact with their GP (FP69) by LA, age, sex
- Estimates by CIS country code and LA
- HESA mode of study by LA
- PR-CIS match rates by age, sex, LA
- Estimates of half weighted records by LA, age and sex
- Patient Register flags by LA or sex or age
- NHS number turnover by LA

In presenting our case studies we make use of these indicators to identify whether the local authorities under study are unusual for these measures in attempt to explain some of the differences observed between SPD V1.0 and official estimates.

Note: These research outputs are NOT official statistics on the population
6.2 Case Study 1: South Bucks

Background

South Bucks is a local authority that lies to the west of London, situated adjacent to the London Borough of Hillingdon. The LA also borders Slough, Windsor and Maidenhead, Wycombe, Chiltern and Three Rivers. Although it lies close to London, it is considerably more rural, categorised under the ONS Urban/Rural classification in 'districts with at least 50 per cent but less than 80 per cent of their population in rural settlements and larger market towns'.

South Bucks had an estimated population of 66,870 in the 2011 Census. This is comprised of small urban areas such as Beaconsfield (population 13,795 in 2011), and smaller towns and villages, many of which are situated on the outskirts of Slough. The area does not contain any major universities, so students are not a significant feature of the population, and neither are armed forces since there are no significant military bases.

South Bucks appears relatively straightforward to estimate in a census compared to many local authorities, which may be connected with the relative absence of certain groups, such as students, that may be challenging to estimate. In the 2011 Census estimates, the overall population estimate had a 95% confidence interval of 0.78%, which is well below the average for all local authorities. However, based on the existing method for producing population estimates from SPD V1.0, a significant proportion of the SPD v1.0 population consists of half-weighted records. In these cases, the matched record is weighted by half in each of the two LAs given on the sources (PR and CIS). The proportion of the population that is made up of these half-weights is generally highest in London. However, South Bucks has the highest level outside London and is higher than several London boroughs. This feature has led to South Bucks being identified as a possible area of interest for investigating the quality of the administrative data.

Comparison of SPD v1.0 and Official Estimates

The total population of South Bucks in SPD v1.0 is 2.8% lower than the official estimates in 2011. Figure 16 shows a comparison of the SPD v1.0 estimates against the census in 2011 and the mid-year estimates in 2013, by single year of age. This shows that there is generally a good relationship between SPD v1.0 and official estimates in 2011, although SPD v1.0 produces lower estimates in some age groups, particularly children, which is also reflected at the national level (see section 4). The main difference seen in 2013 is that SPD v1.0 is higher than the mid-year estimates for ages 20 to 30, whereas in 2011 there is a relatively small difference between the SPD and official estimates.
Figure 16, Comparison of SPD v1.0 and official population estimates by single year of age for 2011 and 2013

SPD v1.0 and 2011 Census population estimates by single year of age
South Bucks, 2011
Total population

Source: Office for National Statistics

SPD v1.0 and 2013 mid-year population estimates by single year of age
South Bucks, 2013
Total population

Source: Office for National Statistics

Note: These research outputs are NOT official statistics on the population
Quality measures for South Bucks

Figure 17 shows the match rate of PR records to CIS in South Bucks in 2011, by single year of age. The match rate is above 90% for nearly all ages and above 95% at many ages. Compared to other local authorities, this is a relatively high match rate between the two sources. For LAs containing significant half-weighted records and/or students, their PR-CIS match rate is often observed to dip around the 20 to 30 age range, much more so than this example for South Bucks. It seems unlikely therefore that the unmatched records on the PR can account for the lower estimates observed on SPD V1.0 compared to official estimates.

In 2011, 10.0% of the SPD v1.0 population of South Bucks consists of half-weights. Slightly more than 1.5% of the population of South Bucks are located there according to HESA data, with the remaining 88.4% of the population consisting of full-weighted records where PR and CIS agree on local authority residence.

South Bucks has the 15th highest level of half-weights among all LAs, and those with higher levels consist entirely of areas in London. The proportion of half-weights is similar in 2013 at 9.6%. For those records with just one of the sources recording local authority residence as

---

8 Match rates have been excluded for 0 to 4 year olds. The SPD V1.0 rule to include all records aged 0 to 4 on the PR effectively rules out matching error as a potential explanation for estimates that differ from official estimates.

Note: These research outputs are NOT official statistics on the population
South Bucks, the aggregate effect of this is to include half the population in South Bucks, and half elsewhere. As this is unlikely to match the true distribution of where people actually live, half-weighting is a cause of uncertainty in SPD v1.0 estimates.

In terms of list cleaning, it has been observed in some LAs that the numbers of FP69s can vary enormously in consecutive years. However, the FP69s are consistently low in South Bucks, with all years from 2011 to 2014 containing fewer than 100 FP69 records. Therefore no significant population changes are expected resulting from exclusion from the SPD from FP69 allocations.

For South Bucks, the levels of internal migration measured by the Patient Register are notable. In the three years from 2011 to 2013, the average annual inflow from other LAs was around 6% of the total number of records on SPD V1.0 in South Bucks. This was approximately balanced by an outflow to other LAs of over 5% each year. This is relatively high compared to many other LAs, with South Bucks ranked between 23rd and 26th out of 348 LAs each year on inflow as a percentage of the SPD V1.0 population. For outflow as a percentage of the SPD V1.0 population, South Bucks ranks between 37th and 46th.

New NHS numbers appear on the PR due to birth, or immigration from abroad, with existing ones disappearing due to death or international emigration. For the SPD V1.0 population around 3% of NHS numbers in South Bucks are new to the PR each year, with a similar proportion of NHS numbers being removed from the PR each year (apart from 2012 when a higher 3.6% of numbers were removed). These are above average but not within the top 25% of all LAs, apart from 2012 when it ranked 40th. This could indicate a higher level of PR list cleaning in 2012 than the years either side.

In summary, the quality measures described above suggest that the main quality issue affecting the SPD in South Bucks is the half-weighted records. There is significant internal migration to and from South Bucks, and this is generally correlated with half-weighted records.

**Half-weights and population churn**

When studying all LAs, a strong correlation is seen between the level of half-weighted records and the level of internal migration shown by the PR. The high level of population turnover in South Bucks suggests that the half-weights are due to one source being updated before the other, after a person has moved.

Figure 18 shows the age and sex distribution of the half-weighted records in South Bucks. This shows that the proportion due to half-weights is far from uniform and that the highest levels are in working age people, particularly males. The highest level is for 30 to 34 year old males, for which the percentage is over 28%, and hence in this group there is a significant uncertainty about how many of these records should in fact be fully attributed to South Bucks. The half-weights follow a similar distribution by age in 2013.
Figure 18, Proportion of the SPD population due to half-weights by five year age and sex in 2011

Figure 19 shows the movement on the PR between South Bucks and other LAs. This shows high levels of inflow and outflow within the range between ages 20 and 45, where the half-weighted records are highest.

Figure 19, Percentage of the SPD population that moves in and out of South Bucks, by single year of age in 2011

Note: These research outputs are NOT official statistics on the population
There is a large spike in Patient Register outflow at age 19, which reaches 31%. These people would have been aged 18 in the previous year when they were last present on the PR in South Bucks. This spike is likely to relate to students moving out of the area to start university. Note that records of students do not result in half-weights as students present on the HESA datasets are fully weighted by their term-time address. Half-weighted records are expected to be due to moves of working age people, with one source being updated before the other.

Figure 20 below shows the prevalence for counting the remaining half weight that is not attributed to South Bucks within other local authorities. The most likely locations for half weight allocations are amongst those neighbouring South Bucks. This suggests that many of the moves are over short distances and it is noteworthy that there are high overlaps across local authorities neighbouring South Bucks that have the same postcode district. For example, it is possible to move between South Bucks and Slough whilst remaining within the same postcode district.

**Figure 20, Location of other half of record for half-weights in South Bucks**

Note: These research outputs are NOT official statistics on the population
This has particular relevance in the case of South Bucks. The matching methodology used to link individual records across datasets partly relies on postcode agreement (either full or partial agreement). As a consequence, a higher proportion of matches are made between records in South Bucks and other local authorities. The impact on the SPD is a higher than average proportion of half-weighted records, despite many of the individuals involved moving only short distances.

Overall the half-weighted records consist of around 53% with the PR record in South Bucks and 47% with the CIS record there. However, half-weights are much more likely to be attributed to the CIS in South Bucks amongst adults under 25. Figure 21 shows that PR half-weights in South Bucks are higher or equal for all five year age groups apart from 15-19 and 20-24 where more are due to the CIS.

**Figure 21, Half-weighted records in South Bucks by whether PR or CIS record indicates South Bucks**

A longitudinal analysis of these records by linking 2011 and 2013 data shows which source is more likely to be correct for location. The half-weighted records in South Bucks in 2011 were linked to the 2013 SPD, and those still present were categorised by whether they were still half-weights, were full-weights in South Bucks, or full-weights in a different LA.

For CIS half-weights in South Bucks in 2011, around 48% of those still present in 2013 had not changed from this state. Of those who had become full-weighed it was much more likely that they were outside South Bucks, with only 14% full-weighed in South Bucks, and 33% elsewhere. Figure 22 shows that this is true across all five-year age groups for the CIS half-weights in South Bucks. However it will have more of an impact on SPD V1.0 estimates for those aged between 20 and 29, as these age groups have the highest proportion of half-weighted records on SPD V1.0.

Note: These research outputs are NOT official statistics on the population
The evidence suggests that the PR is more likely to give the correct location than the CIS. For 2011 half-weighted records with the PR located in South Bucks, around 51% of those remaining in 2013 had not changed from this state. Around 26% became full-weighted in South Bucks, with 18% full-weighted elsewhere.

On balance however, half-weighting is likely to overestimate the population for adult ages up to 40 in South Bucks. This is due to the fact that CIS half-weights are relatively high compared to the PR for younger adults, and these are more likely to have moved out of South Bucks. A revised methodology is needed to more accurately distribute weights across South Bucks and other local authorities, or additional information is needed to determine the correct location of individuals who have conflicting address information. This was discussed as one of the aims of our future research in section 3.6.

Summary

Despite the high level of half-weights in South Bucks there is generally close similarity between SPD V1.0 estimates and the census in 2011. However SPD V1.0 estimates for those aged 20 to 24 are substantially higher than the mid-year estimates in 2013. It is observed that the SPD shows a lower outflow of people at the age where they start university compared to the MYEs. It appears that many students are not updating their PR or CIS when they move to university, and are therefore being put back in South Bucks when they end their studies and are no longer present on the HESA dataset. There is also a high level of uncertainty in moves at the end of study in the mid-year estimates, but more sophisticated adjustments are used. These transitions cause difficulty for both types of estimates so relatively large differences are likely to occur.
6.3 Case Study 2: Hammersmith and Fulham

Background

The Borough of Hammersmith and Fulham is situated in the centre-west of London. The MYE for the population of Hammersmith and Fulham in 2013 was around 179,000, with 25 per cent of its official population aged 25 to 34. Its population is projected to reach 196,000 by 2037. It is the 6th most densely populated local authority in the UK and has the 10th highest proportion of non-UK born residents. Hammersmith and Fulham is characterised by high levels of inward and outward migration both internal and international. These conditions make Hammersmith and Fulham particularly susceptible to high numbers of registrations being recording in the area, which are not always updated when people move out of the area. This case study focuses on the issue of ‘list inflation’, where large increases in the number of administrative records (particularly on the PR) are observed each year.

Comparison of SPD V1.0 with census and mid-year estimates

Figure 23 shows a comparison between the SPD V1.0 population of Hammersmith and Fulham and the 2011 Census. In 2011, SPD V1.0 estimates are considerably lower than census estimates between ages 20-39, predominantly attributable to males. The tendency for SPD V1.0 to produce lower estimates than census estimates is fairly typical amongst London boroughs surrounding Hammersmith and Fulham, particularly for persons aged 20-34 and is discussed in section 5.3.
In 2013 however, SPD V1.0 estimates compare more favourably with the mid-year estimates for all age groups. The gap between 20-39 year olds has been completely closed over the two year period (see figure 24 below).

**Figure 24, Population estimates by five year age group: SPD V1.0 and 2013 MYE**

![Graph showing population estimates by five year age group](image)

In the period 2011 to 2013 the male population of Hammersmith and Fulham increased by around six per cent according to SPD V1.0 while the mid-year estimates showed a decline of around two per cent.

Figure 25 shows the change between 2011 and 2013 for the number of males in Hammersmith and Fulham, for both the MYE and SPD V1.0. Of particular interest is the notable increase for the number of males aged 20 to 26 for SPD V1.0 when compared to the MYE. The most extreme example is for 24 year olds, where the MYE is close to zero change over the period, whereas SPD V1.0 is estimating an increase of 444 persons. For ages between 27 and 37, the MYE are consistently estimating a decrease for the number of males in the area, whereas SPD V1.0 has continued increases for the population in this age group.

**Note:** These research outputs are NOT official statistics on the population.
For females, there are fewer examples of directional differences in population change between the two series. Figure 26 shows that for 20 to 27 year olds, an increase in the MYEs estimate usually corresponds with an increase in the SPD V1.0, however the magnitude of increase is notably higher for SPD V1.0. For ages 28 to 40 the MYEs show notable decreases in female population during this period, which is reasonably well picked up by the change in SPD V1.0, although the level decrease is still not as great as that observed in the MYEs.
Figure 27 below shows a comparison of total populations between the two series for years 2011, 2013 and 2014. In 2011, a difference of over 10,000 was observed between the two series. However an overall decrease of over 4,000 persons between 2011 and 2014 in the MYEs has been met by an overall increase of over 6,000 persons on SPD V1.0. This gives the impression that SPD V1.0 is improving over time. However the converging similarity between the two series is likely to be the result of list inflation on the PR, which could result in overestimation in future years or a sudden reduction in the SPD V1.0 estimate following an intensive list cleaning exercise.

Figure 27, Change in total populations for Hammersmith and Fulham, SPD V1.0 estimates and MYEs

![Graph showing population change over time in SPD v1.0 and official estimates.](image)

Source: Office for National Statistics

Figure 28 below shows how PR estimates have increased in Hammersmith and Fulham over the last seven years leading up to 2014. Following an apparent list clean between 2007 and 2008, where large reductions were seen amongst younger adults, the number of records continues to accumulate year on year. By 2011 it has exceeded the level previously observed before the cleaning exercise in 2007/8 and has gone on to accumulate large numbers of records for those in their late 20s by 2014.\(^9\)

\(^9\) To note that the x-axis on figure 28 relates to age of group in 2014. The start of the peak in 2007 (shown at age 35 in 2014) was actually when this cohort was aged 28.

Note: These research outputs are NOT official statistics on the population
The list cleaning exercise in 2007/8 is likely to have had a lasting impact on the low SPD V1.0 estimate produced in 2011. While the list cleaning exercise will have targeted the removal of records for people no longer present, there will still be a number of missing records on the PR for those who are now resident in the area but haven’t yet registered with a GP. In areas of high inward migration, like Hammersmith and Fulham, these missing records will be significant in number. The gap between the MYE and SPD V1.0 series is in part likely to have been closed over the 2011-2013 period by the accumulation of outdated records for people who have since left the area.

Since 2008 the patient register estimate has steadily increased, suggesting that no major list cleaning exercises have been carried out in subsequent years leading up to 2014. Furthermore, the pattern of change shown by the patient register suggests that a high level of list inflation is building up for some age groups, most notably those aged 30 to 50 and that the PR for Hammersmith and Fulham could be at increasing risk of a substantial list clean. A list clean of the 2007-2008 magnitude would probably result in the SPD V1.0 estimate for males in Hammersmith and Fulham moving sharply downwards and away from the official population estimates at some point in the future.

Given their shared reliance on the patient register it seems likely that both the SPD V1.0 series and the MYEs will underestimate net internal migration into Hammersmith and Fulham for males aged 20 to 29. However, differences in the way international migration is accounted for on the PR and in the MYEs will probably account for a large proportion of the difference in change over time. The mid-year estimates estimated international immigration to Hammersmith and Fulham at about 5,000 persons in 2013. The PR uses a flag called ‘flag4’ to indicate registrations of new migrants in England and Wales. This flag can be used
as a proxy for immigration to local authorities. PR shows that in 2012 and 2013 there were around 6000 to 7000 new flags for new immigrants to Hammersmith and Fulham.

Differences in the way international emigrants are accounted for on the official population estimates and on the PR will also have had an impact. Between 2011 and 2013 the mid-year estimates had an international outflow of 12,600. It is likely that a high proportion of these emigrants did not inform their GP they were emigrating, consequently the PR for Hammersmith and Fulham (and other local authorities) can contain records for international emigrants (as discussed in section 4.1). The methods used to estimate international emigration for the MYEs are currently under review and some of the initial research into improvements suggests that emigration flows from Hammersmith and Fulham in 2011-12 may have been overestimated. The combination of international emigration being overestimated in 2011-12 and the difficulty of accounting for emigration on the patient register probably contribute substantially to the convergence of the SPD V1.0 and MYEs in 2013.

The difficulty in accurately capturing the high level of population change in Hammersmith and Fulham manifests itself in a high number of half-weights. Similar to South Bucks, Hammersmith and Fulham had around 25,600 records with half-weights which is 15 per cent of its total population estimate and is the 4th highest in England and Wales. The high number of half-weights indicates that there is uncertainty regarding the estimate of the resident population in Hammersmith and Fulham as the two sources (PR and CIS) show different information for the same record.

**Summary**

There is evidence that areas of high internal and international migration are vulnerable to a rapid increase in the number of PR registrations over relatively short periods. In some cases, this may have the effect of appearing to improve SPD V1.0 performance, as seems to be the case in Hammersmith and Fulham, where notably lower estimates in 2011 have been corrected for in 2013. It should be noted however that the closing of the gap between SPD V1.0 estimates and official estimates in 2013 is a consequence of directional difference in the measurement of population change. Whereas mid-year estimates have made use of other data sources to identify a decrease in the population of certain age groups, the stock estimate on the PR continues to accumulate over the period. Caution needs to be exercised therefore when interpreting an apparent improvement in SPD V1.0 for 2013, where the evidence available from trends in the administrative sources contradict those found in the mid-year estimates series.
6.4 Case Study 3: Cambridge

Background to the area

Cambridge is a local authority located in the county of Cambridgeshire in the East of England. There are a large number of halls of residence in Cambridge, as the University of Cambridge is located in the LA, along with a number of boarding schools; therefore, students are a significant feature. There are no significant military bases in Cambridge, so armed forces are not a feature of the LA (the total number of armed forces resident in the area in 2013 was approximately 200). In the 2011 Census, Cambridge had an estimated population of 123,867.

SPD V1.0 compared to official estimates

Figure 29 shows SPD V1.0 2013 against the official mid-year estimates (MYE) for 2013 for males in Cambridge by single year of age.

Figure 29, Cambridge SPD V1.0 2013 and MYE 2013, Males

As can be seen from Figure 29, there is a noticeably lower estimate on SPD V1.0 for males between the ages of 18 to 25 when compared to official estimates. Lower estimates are also evident for this age group in 2011 when SPD V1.0 2011 is compared to the census estimates. Similar to the patterns at national level, there is a notably higher estimate on SPD V1.0 for working age males above the age of 30. This is not unique to Cambridge and has therefore not been the subject of detailed analysis in this case study.

Note: These research outputs are NOT official statistics on the population.
Figure 30 shows SPD V1.0 2013 for females at single year of age against the official MYE for 2013.

**Figure 30, Cambridge SPD V1.0 2013 and MYE 2013, Females**

![SPD v1.0 and 2013 mid-year population estimates by single year of age](image)

This shows that there are considerably lower estimates for females between the ages of 18 to 20 on SPD V1.0, when compared to other ages, which was also similar for SPD V1.0 2011. The chart above also shows that there are much higher estimates on SPD V1.0 for females between the ages of 23 to 29, which was not apparent in SPD V1.0 in 2011.

The analysis in this case study has focused on identifying possible explanations for the consistently lower estimates on SPD V1.0 2011 and 2013 for males aged 18 to 25, and females aged 18 to 20.

**Analysis of unmatched HESA records**

SPD V1.0 currently includes HESA records that are linked to either the PR or the CIS. However, there are a number of HESA records that are not linked to the PR or CIS and these are referred to as HESA ‘residuals’. Based on current rules used in the construction of SPD V1.0 these residuals will not be included as they are required to have matched to at least one of these two data sources. In 2011 however, 18.79% of these HESA residuals did in fact link to a census record. This indicates that these particular residuals were actually resident in 2011, and have been incorrectly excluded from the SPD V1.0 population estimate. In Cambridge in 2011, there were 690 HESA residuals (453 males and 237 females) linked to the census. While we anticipate that the majority of students will register...
with GPs in their university area, it is possible that a number of them do not, and will not be registered with a GP elsewhere. Foreign students are perhaps an example of those individuals who are more likely to be excluded on the basis of SPD rules, having not necessarily registered with a GP or applied for a national insurance number if they do not intend to work during their period of study.

Assuming that these HESA residuals matching to census in 2011 should have been included in the population estimate for Cambridge, their inclusion would however have had a very minimal impact on reducing the difference between SPD V1.0 2011 and 2011 Census estimate.

The charts in figure 31 below show SPD V1.0 2011 before and after HESA residuals linked to census are added. SPD V1.0 2011 makes only a marginal difference in bringing SPD V1.0 closer to the official estimates.

Figure 31, SPD V1.0 2011, impact of adding HESA residuals that matched to 2011 Census, males 18 to 25 and females 18 to 20

Source: Office for National Statistics
For both males and females, differences observed between SPD V1.0 and official estimates cannot be attributed to students missing on the SPD known to be resident at the time of census. For 2011 however, there are also HESA residuals that did not link to the 2011 Census. There is more uncertainty as to whether these records were actually in the resident population, as there is no record for them on any of the three data sources; PR, CIS or the census. It is in fact plausible that some of these individuals will have been resident in Cambridge at the time of the census. Student populations are challenging to enumerate when undertaking a census (see Estimation and Adjustment for Communal Establishments Report), and the example given above relating to the possibility that foreign students are disproportionately under-registered on the PR and CIS might equally apply to census enumeration.

We have tested the unrealistic assumption that all of these HESA residuals were in fact resident in Cambridge at the time of the 2011 Census, regardless of whether they linked to a census record or not. This is an unrealistic assumption, as although HESA data is collected annually, there are likely to be a number of records for students studying short term courses (that were over a period not including census day) or exiting their studies early. Figure 32 shows that we still find that the difference between SPD V1.0 2011 estimates and 2011 Census estimates cannot be accounted for by the inclusion of all students recorded in Cambridge on the 2011 HESA dataset.
Figure 32, SPD V1.0 2011, impact of adding all HESA residuals

SPD v1.0 population estimates by single year of age with all HESA residuals added
Cambridge, 2011
Males aged between 18 and 25

Source: Office for National Statistics

SPD v1.0 population estimates by single year of age with all HESA residuals added
Cambridge, 2011
Females aged between 18 and 20

Source: Office for National Statistics

Note: These research outputs are NOT official statistics on the population
For males, the inclusion of all HESA residuals in the SPD V1.0 estimate in 2011 would only have had a notable impact in reducing the difference from census estimates for those aged 20 and 21. It would also have the effect of overestimating for those aged 22 to 25. For females aged 18 to 20 the impact of adding all HESA residuals is mostly observed for 20 year olds, however there would still be a notably lower SPD V1.0 estimate compared to 2011 Census.

Assuming that the coverage of student registrations at Cambridge on HESA data is complete, it would appear from this analysis that the deficit in SPD V1.0 estimates in 2011 for younger adults aged between 18 and 25 cannot necessarily be attributed to student populations. We have therefore extended our analysis to consider the status of PR and CIS records in Cambridge for these age groups in 2011.

**FP69s and half-weights analysis**

One of the risks identified in the use of FP69s as the basis for removal from SPD V1.0, is that it will result in the removal of individuals who neglect to respond to GP correspondence, despite still being resident. An analysis of FP69s for males aged 18 to 25 and females aged 18 to 20 indicates that there has been fairly low and consistent allocation of FP69 flags over the period being investigated. The lower estimates produced by SPD V1.0 do not therefore appear to be the result of list cleaning (see case study for Charnwood, section 6.5).

Half-weights have also been investigated, on the basis that this approach may incorrectly reduce the estimate in Cambridge by apportioning half of the records to the incorrect location (see case study for Charnwood, section 6.5). However, even when re-apportioning all half-weights to be fully counted in Cambridge, once again this has very little impact in closing the gap between SPD V1.0 2011 estimates and 2011 Census estimates for these age groups.

**Analysis of CIS and PR Residuals**

An analysis of CIS and PR residuals (ie. those records not matching between the two sources or to HESA) may provide some indication of where individuals missing on SPD V1.0 may be recorded on the administrative data. It is possible that there has been a failure to match CIS records to the PR, or that some of the non-student population for these age groups have not in fact registered with GPs in the area. We now explore the impact of adding in residuals unmatched records from these two sources to the SPD V1.0 estimate.
By adding the CIS residuals to SPD V1.0 in 2011, the gap between official estimates would have reduced, for both males and females aged between 18 and 21. However, in the case of males, the inclusion of CIS residuals on SPD V1.0 would have resulted in an overestimate of the population for ages 21 to 25. Currently we do not have any supporting evidence that any of these additional records were resident in the Cambridge population at the time of the 2011 Census. It is very likely a proportion of these records will either be erroneous (ie. no longer in the usual resident population for England and Wales) or have moved to another local

Note: These research outputs are NOT official statistics on the population
authority. In practice, and as part of our future research into SPD construction, we will require activity data (see section 3.6) to accurately determine which of these CIS residuals should have been included in the SPD estimate.

The same analysis has been undertaken using PR residuals. Figure 34 shows SPD V1.0 2011 before and after the PR residuals have been added to the SPD V1.0 estimate.

**Figure 34, SPD V1.0 2011 before and after PR residuals have been added**

*SPD v1.0 population estimates by single year of age with PR residuals added*

**Cambridge, 2011**

**Males aged between 18 and 25**

Source: Office for National Statistics

*SPD v1.0 population estimates by single year of age with PR residuals added*

**Cambridge, 2011**

**Females aged between 18 and 20**

Source: Office for National Statistics

Note: These research outputs are NOT official statistics on the population
For males, the inclusion of all PR residuals on SPD V1.0 in 2011 would have resulted in notably higher estimates than the 2011 Census estimate. This reflects the general tendency for GP registrations to accumulate for younger adult males as a result of delays in re-registration when moving out of the area (see the case study for Hammersmith and Fulham at section 6.3).

For females, where there is typically less delay in GP registrations, the inclusion of all PR residuals on SPD V1.0 in 2011 also overestimates when compared to 2011 Census estimates for these age groups. The fact that residual inclusions from the PR would have overestimated the population for both males and females might be indicative that some of these residual records should have been included in SPD V1.0, but have been excluded on the basis that they have not matched to the CIS.

Similarly to the CIS, the basis for including PR residuals on the SPD requires further evidence of recent activity to confirm residency in the Cambridge area. Examples of this would include recent interaction with health services, such as GP or hospital appointments. While the inclusion of all PR residuals for males and females in Cambridge in 2011 would have produced a higher SPD V1.0 estimate than the 2011 Census estimate, activity data could be used as the basis for detecting and removing erroneous records or individuals who are likely to have left the area.

**Match rates between PR and CIS**

Something also of particular interest in Cambridge is the relatively lower match rate observed between these two sources for both males and females within the age groups of interest. Figure 35 shows the match rate between the PR and the CIS for 2011 data for males and females. Both charts demonstrate that Cambridge has a much lower match rate than the national average and even below average for London Boroughs, where match rates are typically lowest in the country.
Figure 35, Cambridge match rates compared with national and London average

Further analysis is required to understand the reasons behind the significantly lower match rate in Cambridge for these age groups. It may be the case that there has been a failure to match records between the PR and the CIS in these areas, which has resulted in a large

Note: These research outputs are NOT official statistics on the population
number of exclusions from SPD V1.0. The development of matching algorithms to include a probabilistic stage in an attempt to link some of the residual records is currently being developed. This may have the effect of identifying more true matches between sources, which will in turn increase the SPD V1.0 estimates in Cambridge for these age groups. It should also be noted that this is also expected to have a positive impact in London Boroughs where match rates for the same age groups are typically low.

Summary

This case study set out to explain possible reasons for lower estimates on SPD V1.0 compared to official estimates for males aged 18 to 25 and females aged 18 to 20 in Cambridge. Having tested the impact of adding in residual HESA records that are unmatched to both the PR and CIS, it would appear that the lower estimates observed in Cambridge are not necessarily the result of missing student records that have been excluded from SPD V1.0. The exceptionally low match rates between the PR and the CIS for the age groups under study may imply that some Cambridge residents are only recorded on one of these data sources, or that we have been unsuccessful in matching individuals’ records to include them in the SPD V1.0 population estimate. Improved matching algorithms and the use of additional data sources (including activity data), may in future enable us to identify and include additional records that have been incorrectly excluded from SPD V1.0 based on the current methodology.
6.5 Case Study 4: Charnwood

Background to the area

Charnwood is a local authority in the East Midlands, situated to the north of Leicester. Its largest urban area is Loughborough, which had a population of 59,932 in the 2011 Census, and it has a significant student population of around 16,000 at Loughborough University.

International migration is relatively low for Charnwood. It had an inflow of about 1,500 people from abroad in 2012 and 2013, and an outflow to countries abroad of about 500 people. In comparison the internal migration to and from other areas in the UK is nearly ten times this, with an inflow of between 10 and 11 thousand people each year. Although the net population change through internal migration is modest, the large gross inflow and outflow of individuals moving between areas is a particular challenge for estimating the population in Charnwood.

Charnwood has been identified as an area worthy of further study because it has a large student population and much of the population change in the area is through internal migration of students or people of student age. This case study compares SPD v1.0 and the official population estimates to identify discrepancies that deserve investigation, then uses the quality indicators (outlined in 6.1) to investigate possible causes of discrepancies.

Comparison of SPD v1.0 and official estimates

Figure 36 shows the population by single year of age in Charnwood, using both official and SPD v1.0 estimates in 2011. It can be seen that for each of the years, the two estimates are generally close together. Figure 37 shows the population in 2013 and appears similar except from the ages between 18 and 22 where SPD v1.0 is significantly lower than official estimates in 2013. The total difference between SPD v1.0 and the official estimates at ages 18 to 22 in 2013 was 2,300 people. In 2011 there was a difference of only 500 people.
Figure 36: Comparison of SPD v1.0 and official population estimates by single year of age for 2011

Figure 37, Comparison of SPD v1.0 and official population estimates by single year of age for 2013

Note: These research outputs are NOT official statistics on the population
The lower SPD v1.0 population in 2013 at student ages is a population change that does not corroborate with the official estimates. This change therefore appears suspicious and may be due to effects taking place within the administrative data. The following sections examine the quality indicators to attempt to understand this change in the SPD.

**Quality measures for Charnwood**

This section looks at the various quality measures about the administrative data to understand possible causes of the discrepancies described above. Figure 38 shows the match rate to CIS for Patient Register (PR) records in Charnwood in 2011, by single year of age. The lowest rate corresponds with people aged 23 and 24, however the match rates shown are broadly similar to other local authorities with large student populations. Figure 39 shows the match rates for 2013, which are very similar. The similarity between years suggests that the anomaly seen in figure 37 of lower student estimates in SPD v1.0 does not relate to the matching process. Note that students may also be included in SPD v1.0 via matches between Higher Education Statistics Agency (HESA) data and either of the PR or CIS, but the number of unmatched HESA records is low in both 2011 and 2013. Therefore unmatched student records do not contribute to the difference seen between years in figures 36 and 37.

**Figure 38, Rate of matching of PR records to CIS for males and females in Charnwood in 2011, by single year of age**

![Proportion of Patient Register records matched to CIS by single year of age in Charnwood, 2011](image)

**Source:** Office for National Statistics
The proportion of records that are half-weighted to Charnwood in SPD v1.0 is 5.2% which is close to the median (5.1%) amongst all local authorities. Some 8.6% of records counted in Charnwood are from students allocated via their HESA term-time location. The remaining 86.2% records are fully weighted in Charnwood indicating a good degree of agreement between the PR and CIS, and relatively low uncertainty about the location of residents registered in Charnwood on the administrative data.

Year-to-year changes in the Patient Register affecting a local authority are categorised into different types of change or turnover and are shown in Figure 40. The different types are explained earlier in this document. Note that when looking at a single area, the change in NHS numbers can be broken down into those individuals moving into an area and those moving out to elsewhere in the UK. It can be seen from Figure 40 that migration of people to and from other LAs dominates the picture for Charnwood. Broadly speaking this agrees with the mid-year estimate components of population change. For 2011-12, the PR change implies similar change due to births deaths and international migration to official estimates, but lower movements of people between Charnwood and other areas. For 2012-13, the Patient Register has a lower change due to births and international migration than the official estimates and records lower movements of people between Charnwood and other areas. The net effect of people migrating between areas appears to be more positive in the Patient Register though, most likely due to migrations out of Charnwood being less well captured.
Figure 40, population change in the Patient Register

Year to year changes on the Patient Register
Charnwood, 2010 to 2013

Source: Office for National Statistics

Figure 41 shows the changes to the Patient Register for 18 to 22 year olds only for the time periods 2011-12 and 2012-13. The changes are quite consistent comparing the periods, and we can see that movements between Charnwood and other LAs dominate this age group. This is a similar pattern to the components of change for official population estimates but with lower gross flows. The relative consistency of the Patient Register changes for this age group suggests that although the PR may record less change than official estimates the raw volume of change is unlikely to drive the difference we see in SPD v1.0 in 2013. It should however be noted that not all Patient Register records are included in SPD v1.0. Unmatched records are excluded as well as records with an FP69 flag.
Figure 41, population change in the Patient Register ages 18-22 only

Year to year changes on the Patient Register for persons aged 18 to 22

Charnwood, 2011 to 2013
Persons aged 18 to 22 at the end of each time period

Figure 42 shows the occurrence of records with an FP69 flag in the Patient Register for males in 2011, 2013 and 2014. It can be seen that similar low numbers of records with this flag occur in 2011 and 2012. The number of FP69s for females (not shown) is similarly low in 2011 and 2012. In 2013, the occurrence of FP69s for males is considerably higher, and this increase is similarly notable amongst females (also not shown). Such a large increase in the number of FP69 allocations has led to a much higher number of exclusions from SPD v1.0 in 2013. The higher occurrence of FP69s in 2013, particularly around student ages is the main explanation for the difference between SPD v1.0 and official estimates seen in 2013.

Note: These research outputs are NOT official statistics on the population
In summary the quality indicators show that Charnwood does not lie any great distance from the average of the measures for all LAs, and appears fairly consistent over time for the ages 18 to 22. The number of FP69s is the only measure where there appears to be an unusual increase in 2013.

Since there is uncertainty over whether people with an FP69 flag are still present in the LA, the SPD method excludes them from the population. However, authorities have been observed to make inconsistent use of FP69s, issuing very few in some years, but occasionally a year occurs in which a significant proportion of a particular LA are issued an FP69.

In 2011, there were 179 PR records in Charnwood with an FP69, close to the median occurrence of FP69s for local authorities. In 2013 there were 3,060 FP69 records in Charnwood, the 15th highest total of that year. This increase of 3,030 between 2011 and 2013 is the 11th largest increase, with neighbouring Leicester having the 4th largest increase. It appears that the health authority identified more list-cleaning candidates in 2013 than in 2011. Hence the number of FP69s may not be a reliable measure of records needing to be removed and not a sound way in which to identify non-usually resident people for exclusion from SPD v1.0. It appears to measure the level of effort made by the authorities at the time. Indeed, further analysis of FP69s shows that of those in Charnwood in 2013, 1,004 were still on the PR in Charnwood in 2014, with 600 in the PR in a different LA. Therefore a significant fraction of FP69s are not actually removed from the list, which also raises doubts about the value of using FP69s in creating the SPD population.

Note: These research outputs are NOT official statistics on the population
Further analysis shows that the FP69s in Charnwood mostly occur in the young adult age group, where the discrepancy is seen in Figure 42. For males aged 20 to 24, there are 32 FP69s in 2011 and 1,285 in 2013. This increase of 1,253 is the largest increase among all LAs, and this age group comprises around 40% of the FP69s in Charnwood in 2013. When comparing the 2013 mid-year estimates to the 2013 SPD v1.0 for all people aged 20 to 24, the SPD v1.0 estimate is 1,220 lower than the MYE population for Charnwood. Therefore, including the additional FP69s that occurred in 2013 for males aged 20 to 24 would be sufficient to bring the population for all 20 to 24 year olds back into line with the official estimates.

**Summary**

Analysis of the quality indicators has shown that the population change observed in Charnwood between 2011 and 2013 is largely due to operational processes involved in the collection of administrative data. The large increase in the assignment of FP69s by the health authority has had significant consequences for SPD v1.0. This highlights potential issues with the use of FP69s in filtering records out of the SPD. Charnwood appears similar to most LAs in that FP69s are low in most years, with occasional increases to much higher numbers. This suggests that more list cleaning effort is made in some years than others, and hence the SPDs in years where lots of effort is made may have a significant level of wrongly excluded records, whilst other years may have some records wrongly included. These changes can cause apparent volatility in the SPD V1.0 population estimates, and hence it may be advantageous to consider changing the current SPD methodology in this respect in the future.

Note: These research outputs are NOT official statistics on the population.
7.0 Conclusions and next steps

The estimates produced from the SPD V1.0 method shows considerable promise for estimating the population. For the majority of local authorities, the SPD V1.0 estimates are similar to the official estimates for all three years included in this release; 2011, 2013 and 2014. As highlighted in this report, there are a small number of local authorities, largely in London and other city areas, where the SPD V1.0 estimates differ notably from official estimates. This highlights the need for access to more administrative data, and that a Population Coverage Survey may be needed to assess and adjust for coverage errors in SPDs.

Future releases of SPD population estimates will continue to focus on improving the coverage of particular age-sex groups at local authority level incorporating the use of other datasets. In future, we will aim to include the English and Welsh School Census in our SPD methodology to improve the coverage of children of school age, as well as the use of activity data (subject to suitable data access) to identify erroneous records that have contributed to an overestimate in the number of working age males.

Of crucial importance is the need to develop new methods for independently evaluating the SPD population estimates. Developing this method remains an important part of our future research.

Comparisons with the mid-year population estimates and the census have highlighted differences that need further analysis. We will be undertaking further research to understand these discrepancies and consider the potential implications for developing the existing methodology for producing existing official population estimates. We are keen to work with users of the data to further understand possible explanations for differences that arise between the SPD series and the official estimates. As part of the user feedback process we would be interested to hear about methods that could be used to independently evaluate the SPD without reliance on official estimates.

We will continue to work with our data suppliers, feeding back our analysis and research findings with the aim of improving the quality of administrative data for the purpose of producing population statistics. This will be an ongoing process.

As outlined in the introduction of this report, the administrative data research outputs will be produced on an annual basis. There will be an annual cycle to support this process, including engagement with users to get feedback, as well as developing methods and the range of outputs. In future years we are aiming to release outputs for population estimates by single year of age and at smaller geographic levels, as well as outputs for new topics, which may include estimates based on qualifications, personal or household income, and ethnicity, as well as on the number and the size and composition of households.

Note: These research outputs are NOT official statistics on the population
# Annex A – Glossary

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity data</td>
<td>Information from administrative data sources about when individuals have interacted with systems or services, such as the National Insurance, tax or benefits systems, or a hospital visit through the NHS system.</td>
</tr>
<tr>
<td>Administrative data</td>
<td>This is data that people have already provided to government, for example in the course of accessing public services. Some of this data could be re-used by ONS to produce statistics about the population.</td>
</tr>
<tr>
<td>CIS</td>
<td>Customer Information System of the Department for Work and Pensions. An administrative data source which contains a list of people who have a National Insurance Number.</td>
</tr>
<tr>
<td>Confidence interval</td>
<td>A 95 per cent confidence interval is a range within which the true population would fall for 95 per cent of the times the sample survey was repeated. For example, for a 95% confidence interval, the true (unknown) value of the estimate would be expected to lie within it 19 times out of 20. A more detailed explanation is provided in the Paper O4.</td>
</tr>
<tr>
<td>DWP</td>
<td>Department for Work and Pensions</td>
</tr>
<tr>
<td>Flows, flow based approach</td>
<td>This refers to population change between two time points, due to births, deaths and movements of people through internal and international migration.</td>
</tr>
</tbody>
</table>
FP69 flag
This is an indicator that there has been no contact between a GP practice and an individual in the last 15 months. An attempt to contact at the address has been made by the GP practice but no communication has been received. It cannot be assumed that the individual still resides at the address (these records are excluded from SPD v1.0).

HESA
Higher Education Statistics Agency whose data contains a list of students who are registered on to a higher education course in England and Wales.

HMRC
Her Majesty’s Revenue and Customs

Internal migration
This refers to residential moves between different geographic areas within the UK. This may be between local authorities, regions or countries.

International migration
ONS uses the UN recommended definition of a long-term international migrant: “A person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of destination effectively becomes his or her new country of usual residence.”

Local authority (LA)
Total population between 2,200 and 1,074,000 people, average 160,000 people.

Lower layer super output area (LSOA)
Total population between 1,000 and 3,000 people, average 1,600 people.

Middle layer super output area (MSOA)
Total population between 5,000 and 15,000 people, average 7,800 people.

MYEs
Mid-year official population estimates

10 For an introduction to the different types of geography see http://www.ons.gov.uk/ons/guidemethod/geography/beginner-s-guide/index.html
Net migration

Net migration is the difference between people moving into an area and people moving out of the same area. If net migration is positive then it means that more people have moved to live in the area than have left to live elsewhere.

NHS

National Health Service

Output area (OA)\(^{11}\)

This is the lowest geographical level at which census estimates are provided. Total population between 100 and 625 people, average 300 people

Official estimates

In this document, official estimates refer to population statistics from the 2011 Census and annual mid-year population estimates (MYEs).

P1 quality standard

Maximum quality standard set out in the evaluation criteria of the Beyond 2011 programme. This is equivalent to the maximum quality achieved in the current system (ie what is achieved in a census year), every year (see Beyond 2011 Options Report 2 for more information).

P3 quality standard

Average quality standard set out in the evaluation criteria of the Beyond 2011 programme. This is equivalent to the average quality achieved in the current system, every year (see Beyond 2011 Options Report 2 for more information). for more information).

Population Coverage Survey (PCS)

The aim of the PCS is to facilitate the estimation of the population at LA level by age and sex, by making adjustments to the population totals provided by the SPDs. The current assumption for the PCS design is that it would cover approximately 1% of the population in England and Wales, or around 350,000 households. The development of the proposed design is described fully in Estimating the Population: In Theory.

\(^{11}\) For an introduction to the different types of geography see http://www.ons.gov.uk/ons/guidemethod/geography/beginner-s-guide/index.html

Note: These research outputs are NOT official statistics on the population
Pseudonymisation

Pseudonymisation is a procedure by which identifying fields (i.e. names, dates of birth and addresses) within a data record are replaced by one or more artificial identifiers to protect the privacy of individuals.

Statistical Population Dataset (SPD)

Once individual records have been matched across data sources using the approach outlined in *Estimating the Population: In Practice*, the information is pulled together into a single, coherent dataset that forms the basis for estimating the population. This is called a Statistical Population Dataset (SPD).

Stock estimates or Stock based approach

Refers to a snapshot of the population at a point in time.

Working age population

People aged 16 to 64.

Note: These research outputs are NOT official statistics on the population