

Article

Developing our approach for producing admin-based population estimates, England and Wales: 2011 and 2016

Research into developing a new methodology to create population estimates from administrative data. These estimates are not official statistics on the population.

Contact:
Becky Tinsley
pop.info@ons.gov.uk
+44 1389 444661

Release date:
21 June 2019

Next release:
To be announced

Table of contents

1. [Disclaimer](#)
2. [Main points](#)
3. [Things you need to know about this release](#)
4. [Introduction](#)
5. [Using activity data to reduce overcoverage](#)
6. [Principles of ABPE V3.0 methodology](#)
7. [Development of the rules for ABPE V3.0](#)
8. [How does the new method compare with official estimates?](#)
9. [Summary and next steps](#)
10. [More about our transformation journey](#)
11. [More about our official population and migration statistics](#)

1 . Disclaimer

These Research Outputs are not official statistics on the population nor are they used in the underlying methods or assumptions in the production of official statistics. Rather, they are published as outputs from research into a methodology different to that currently used in the production of population and migration statistics. These outputs should not be used for policy- or decision-making.

2 . Main points

We have continued the research outlined in our previous [slide pack](#), using activity-based rules and records from single and linked data sources to develop our approach for producing admin-based population estimates. Initial rules have been combined to produce the first admin-based population estimates based on this approach, which have been produced for years 2011 and 2016. These admin-based population estimates for England and Wales have been compared with official estimates for these years.

These rules have largely removed patterns of overcoverage seen in previous estimates from administrative data. While this now results in higher levels of undercoverage, this is much more comparable with the results seen from the census before any coverage adjustment takes place.

There is potential for this method to be improved to reduce the amount of coverage adjustment required. Access to additional data sources and more sophisticated rules are likely to allow for more records to be included where there are currently deficiencies, without increasing the levels of overcoverage. This means any adjustments will be from a better starting point than the census.

This provides a platform for combining with a Population Coverage Survey to produce coverage-adjusted estimates in a similar way to the method used to successfully adjust the census.

3 . Things you need to know about this release

We are transforming the way we produce population and migration statistics to better meet the needs of our users and to produce the best statistics from the best-available data. For information on this transformation see [our overview of the transformation of the population and migration statistics system](#).

The analysis in this report advances the previous research we have undertaken to produce estimates on the size of the population using administrative data, previously known as a [Statistical Population Dataset](#) (SPD). Throughout this report we will refer to this approach as admin-based population estimates (ABPE).

This report shares initial results for the first attempt at a new approach for producing admin-based population estimates using activity-based rules at the national level. We recognise that more work is required to refine and develop the methodology. Further work is also required to understand the impact of using this method for local-level estimates.

As in our previous research, our methodology is based on anonymously linking person records on administrative datasets to construct administrative-based population estimates. For information about our previous methodologies, please see [SPD V1.0](#) and [SPD V2.0](#) methodology reports.

For further information on the data sources included in these Research Outputs see the [data source overviews](#).

We welcome users providing feedback to us on their quality, value or the impact that using these figures would have if they were used in place of official statistics, please contact us pop.info@ons.gov.uk.

4 . Introduction

We have previously produced [research-based population estimates](#) using administrative data by linking pseudonymised records between multiple datasets and applying a simple set of rules in an attempt to replicate the usually resident population¹.

For the previous method to produce [admin-based population estimates \(ABPE\)](#) (previously Statistical Population Dataset (SPD) V2.0), four data sources were used:

- the NHS Patient Register (PR)
- the Department for Work and Pensions (DWP) Customer Information System (CIS)
- data from the Higher Education Statistics Agency (HESA)
- School Census data

Records found on two of these four data sources were included in the population. This [resulted](#) in [estimates higher than the official estimates](#) (or “overcoverage”) for certain population groups, especially males of working age. This is likely to be because of people leaving the country but not being removed from the administrative sources.

In our previous publications, we have explained the intention to combine the ABPE with a Population Coverage Survey (PCS). The PCS would be designed to measure coverage errors seen in the ABPE. The ABPE and PCS could be combined and used within an estimation method to produce robust, high-quality population estimates – in much the same way that the Census Coverage Survey adjusts the census results. However, earlier work showed that [existing estimation methods](#) (such as dual system estimation) struggle to produce robust estimates when the underlying data contain high levels of overcoverage.

The objective of the new ABPE method, ABPE V3.0, is to use new data sources to remove records that are erroneously included in previous methods. An ABPE with undercoverage for all age and sex groups would be more similar to unadjusted census responses and should allow dual-system type estimators to be applied with improved results. The new method therefore sets relatively strict criteria for inclusion in the population.

A comparison between ABPE V3.0 and unadjusted 2011 Census counts is made in [Section 8](#) of this report. Future improvements to the ABPE could then gradually add more records, reducing the size of adjustment needed by estimation, but only if this does not result in erroneous records being included.

Notes for: Introduction

1. We are currently adopting the UN definition of “usually resident” – that is, the place at which a person has lived continuously for at least 12 months, not including temporary absences for holidays or work assignments, or intends to live for at least 12 months (United Nations, 2008).

5 . Using activity data to reduce overcoverage

In our previous Research Outputs, we conducted some initial research using “activity”¹ data from income and benefits sources to [remove records for those aged 16 years and over with no recent interactions](#) on these administrative systems. This resulted in around 10% of records being removed for ages ranging from the mid-twenties up to State Pension age, with much higher reduction in population for younger ages. Therefore, an activity-based approach showed potential to remove the overcoverage, producing admin-based population estimates (ABPE) to which a coverage adjustment could more easily be applied.

Another potential issue with the previous method is that new migrants may appear on only one source for some time after their arrival, meaning these records would be excluded from the ABPE population, limiting the effectiveness of the method to measure international migration. Our research has also shown that there can be a [delay in new migrants registering for more than one service](#), providing further evidence for the need to review the ABPE methodology. With the availability of high-quality activity data, evidence from a single source should be sufficient to indicate presence in the population.

Developing on this, the ABPE V3.0 method requires evidence of activity for inclusion in the population and includes records from both single and linked data sources. This report shares the methodology and illustrates strengths and weaknesses of this approach using analysis at the national level in comparison with the old approach and with official population estimates. Further information on the methodology implemented can be found in [Section 7](#).

Notes for: Using activity data to reduce overcoverage

1. "Activity" can be defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way. Only demographic information (such as name, date of birth and address) and dates of interaction are needed from such data sources to improve the coverage of our population estimates.

6 . Principles of ABPE V3.0 methodology

We have taken a data-driven approach to building our rules, [identifying the data sources that provide the best coverage for a given age group](#). Simple rules for each group are combined to produce the full admin-based population estimates V3.0 (ABPE V3.0), with all records included following the two basic principles:

- all records included should have a sign of activity within the 12 months prior to the reference date of the ABPE (for example, 31 March 2011: Census date, 30 June 2016: official population estimate reference date), or appear in the same address and have a relationship with an active person
- activity and appearance on a single data source is sufficient for inclusion in ABPE V3.0, with data linkage only necessary to deduplicate records that appear on multiple sources

Since official population estimates are produced annually, activity within the previous 12 months is the minimum requirement for ABPE V3.0. However, this does not indicate that a record has evidence that it is still present at the reference point. Additionally, ABPE V3.0 methodology requires that the records refer to a stay in the UK that is longer than 12 months, meeting the definition of a usual resident population. As a result, the ABPE is still likely to contain some records that would not appear in the official estimates.

On some data sources, such as Higher Education Statistics Agency and School Census, registration takes place annually, providing little evidence of activity at other times of the year. Therefore, the annual registration is taken as the source of activity and so these sources only satisfy the previously described minimum requirement. This means that while there is evidence of activity within the 12 months prior to the reference date, further evidence is required to confirm residence for 12 months or longer.

In contrast, benefits data give information on dates of claims to the nearest month. Here more strict rules can be applied, including only those claiming in the reference month. This will provide more confidence that the record can be defined as usually resident on the reference date. Again, further evidence is required to confirm residence for 12 months or longer.

Although the method uses data sources individually, records with activity on multiple sources will appear as duplicates. To prevent this, data sources are linked to ensure a record is only included once in the population.

Summary of ABPE V3.0 rules

Table 1 shows a summary of the data sources and rules that are used to determine records included in ABPE V3.0. The data sources and development of these rules are described in more detail in the following section.

Table 1: Data sources and rules used in ABPE V3.0

Source of activity	Inclusion rules
Pay As You Earn (PAYE) and Tax Credits data	<p>For 2011 estimates include all people with a positive amount earned on either PAYE or Tax Credits in the tax year ending 2011.</p> <p>For non-Census year estimates (at mid-year), include those with a positive amount earned on either PAYE or Tax Credits in both the current and previous tax years.</p>
National Benefits Database (NBD), Housing Benefit (SHBE) data	Records showing a claim in the same month as the ABPE V3.0 reference date are included.
Child Benefit data	Records with age 0 to 15 years are included if the claim is in the same month as the ABPE V3.0 reference date. Records with age 16 to 19 years are included if the claim is within the previous 12 months of the reference date.
NHS Patient Register (PR) and Personal Demographic Service (PDS) data	<p>PDS is used if available for the relevant year, otherwise PR is used.</p> <p>For PDS, include records if the initial registration is within the 12 months prior to the ABPE V3.0 reference date, or if the latest assignment to a GP service is within this period.</p> <p>For PR, include records if the initial registration is within the 12 months prior to the ABPE V3.0 reference date.</p>
Higher Education Statistics Agency (HESA) data	<p>Include all UK domiciled student records, excluding records identified as dormant or studying abroad for part of the degree.</p> <p>Non-UK student record included if identified as studying on a course for at least one year.</p>
English and Welsh School Census data	Include all records from the English and Welsh School Census extracts taken in the ABPE reference year.
Births registrations data	Include all records for children who were born in England and Wales within the 12 months prior to the ABPE V3.0 reference date.
Inactive records	<p>Include records showing no activity that match with a surname of the record with the highest age of an active record in the same address.</p> <p>Exclude those records that are more than 16 years older than the active record, or if the inactive record is identified as aged 18 years.</p>

7 . Development of the rules for ABPE V3.0

This section describes the individual data sources used in the creation of admin-based population estimates V3.0 (ABPE V3.0) and rules for each of them. The aim of these initial rules is to have relatively strict criteria for inclusion, so we can be more confident that the records included in the new ABPE meet the usually resident definition. As a result, this is likely to be lower than a “true” count of the resident population.

Benefits and Income Datasets (BIDS)

The BIDS data used in ABPE V3.0 comprise Pay as You Earn (PAYE), Tax Credits, National Benefits Database (NBD), Single Housing Benefit Extract (SHBE) and Child Benefit data.

This combination of sources covers the majority of types of income, the major exception being those who only have self-employment or other rarer sources of income. Therefore, these datasets are expected to be the main source of activity for a large majority of the population and are fundamental to the ABPE V3.0 method. For further information on the coverage of these data sources see the [data source overview on income and benefits](#).

Of these datasets, PAYE is the largest, including information about individuals in employment (excluding those in self-employment) and individuals receiving an income. The dataset is an annualised extract containing one record per tax year per individual. Therefore, indicating presence in the population for part of the tax year. The Tax Credits dataset is combined with the PAYE data. Activity in a tax year can come from either source, with an income greater than zero on either being considered a sufficient indicator.

The 2011 ABPE V3.0 refers to the end of March 2011, to match the census estimates. This aligns with the end of the tax year ending 2011, so all records with activity in this tax year are included.

Just over 30 million records are contributed to by the PAYE data. The Tax Credit data provide an additional 3 million records approximately. Initial analysis of the data shows that the majority of Tax Credit claimants are female.

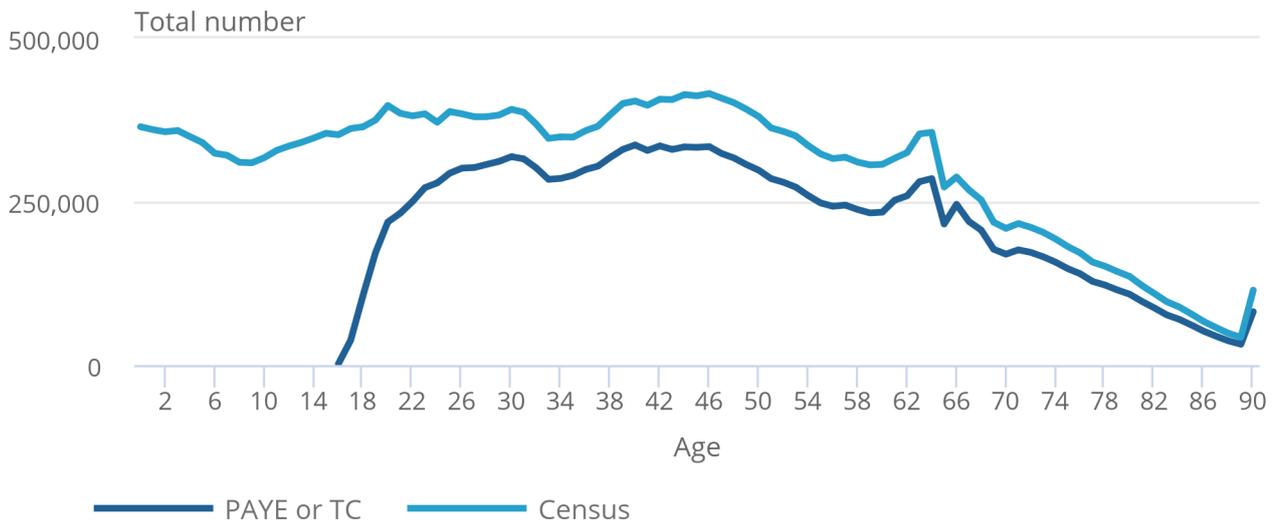
ABPE V3.0 has also been produced for 2016 and this version is designed for comparability with the official mid-year estimates. In this case, activity is required on both the current and previous tax years to minimise overcoverage. Records appearing only on the previous tax year may have emigrated and those appearing only on the current tax year may not be usual residents.

Figure 1a: ABPE V3.0 population with activity on Pay As You Earn or Tax Credits datasets, compared with census estimates

England and Wales, 2011, males

Figure 1a: ABPE V3.0 population with activity on Pay As You Earn or Tax Credits datasets, compared with census estimates

England and Wales, 2011, males



Source: Office for National Statistics

Notes:

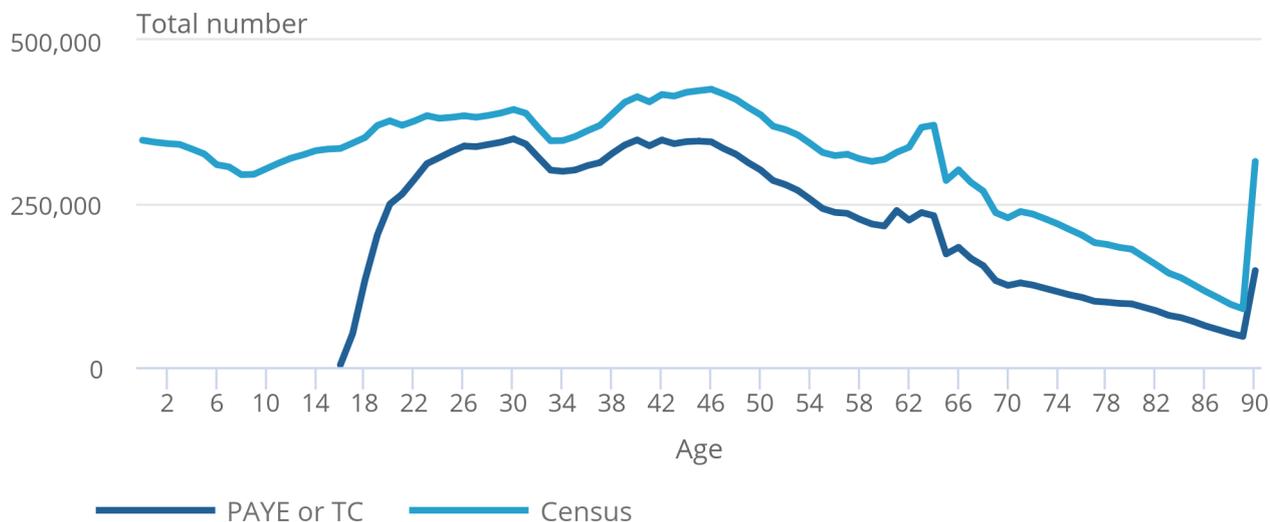
1. ABPE – Admin-based population estimates.
2. "Activity" is defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.
3. PAYE – Pay As You Earn.
4. TC – Tax Credits.

Figure 1b: ABPE V3.0 population with activity on Pay As You Earn or Tax Credits datasets, compared with census estimates

England and Wales, 2011, females

Figure 1b: ABPE V3.0 population with activity on Pay As You Earn or Tax Credits datasets, compared with census estimates

England and Wales, 2011, females



Source: Office for National Statistics

Notes:

1. ABPE – Admin-based population estimates.
2. "Activity" is defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.
3. PAYE – Pay As You Earn.
4. TC – Tax Credits.

Figures 1a and 1b show that the combination of PAYE and Tax Credits data includes a large majority of the adult population as expected. Some differences between males and females are seen, particularly for pensioners where the numbers for males are closer to the census estimates than for females. For this age group it is likely that more males have occupational pensions than females so are more likely to appear on the PAYE data.

For those aged in their twenties and early thirties, numbers of females are closer to census estimates than males. This may be because of the Tax Credits data being female-dominated, as well as missing records for those who are self-employed, of which males are expected to be a majority, as indicated by [2011 Census data](#).

In contrast to PAYE and Tax Credits, the benefits datasets of NBD, SHBE and Child Benefit include the month of the beginning and end of each claim, giving more precise dates for activity.

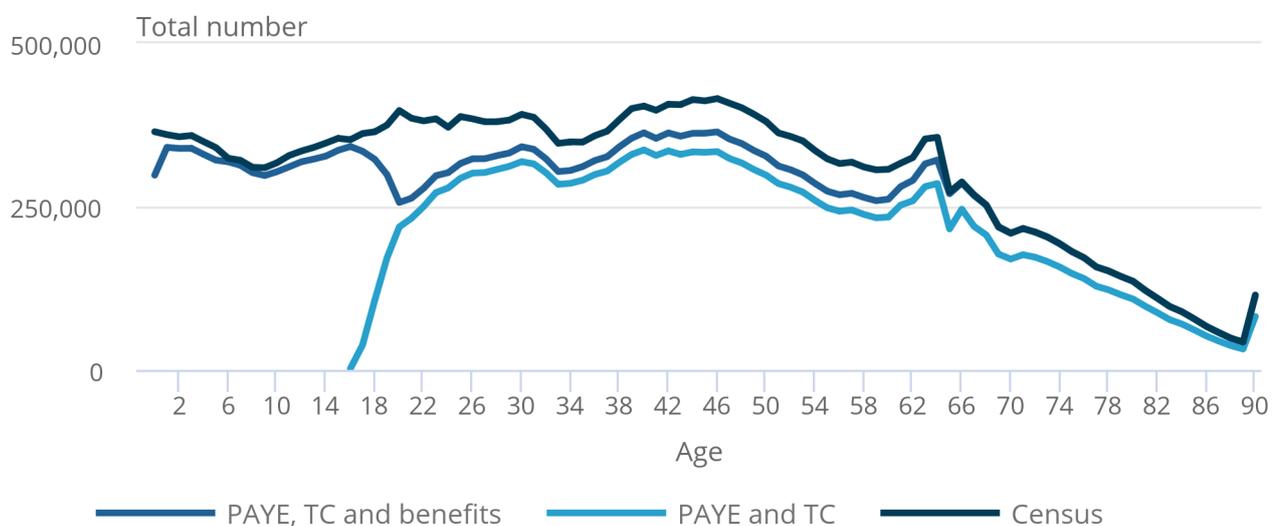
To minimise the possibility of overcoverage, only records with a claim date in the same month as the ABPE reference date are included, except for some Child Benefit records. Child Benefit eligibility stops at age 16 years if the child is no longer in education, but otherwise may continue up to age 19 years. In this age range claims are likely to end and it may be some time before records gain activity on other data sources. For this age range it is sufficient to have had a Child Benefit claim within the previous 12 months of the reference period; without this exception the numbers in the ABPE would be substantially reduced for this group.

Figure 2a: ABPE V3.0 population after adding other active benefits records, compared with census estimates

England and Wales, 2011, males

Figure 2a: ABPE V3.0 population after adding other active benefits records, compared with census estimates

England and Wales, 2011, males



Source: Office for National Statistics

Notes:

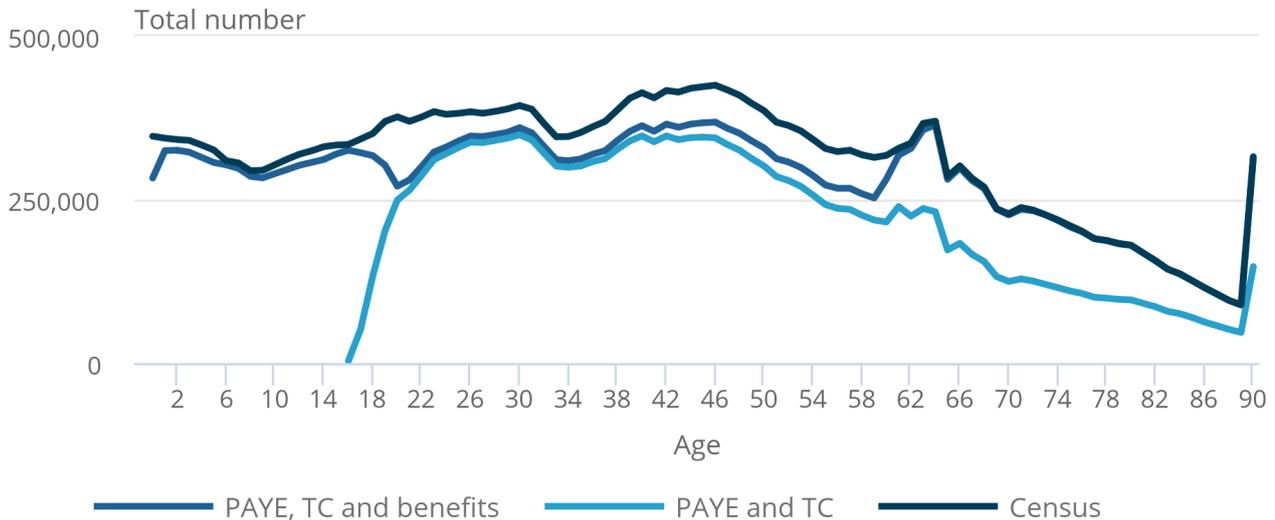
1. ABPE – Admin-based population estimates.
2. “Activity” is defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.
3. PAYE – Pay As You Earn.
4. TC – Tax Credits.
5. Other benefits – these include benefits from the National Benefits Database (Incapacity Benefit, Jobseeker’s Allowance, Carer’s Allowance, Income Support, Severe Disablement Allowance, Employment and Support Allowance, Widow’s Benefit, Bereavement Benefit, Disability Living Allowance, Pension Credit, State Pensions and Attendance Allowance) and records from the Single Housing Benefit Extract.

Figure 2b: ABPE V3.0 population after adding other active benefits records, compared with census estimates

England and Wales, 2011, females

Figure 2b: ABPE V3.0 population after adding other active benefits records, compared with census estimates

England and Wales, 2011, females



Source: Office for National Statistics

Notes:

1. ABPE – Admin-based population estimates.
2. "Activity" is defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.
3. PAYE – Pay As You Earn.
4. TC – Tax Credits.
5. Other benefits – these include benefits from the National Benefits Database (Incapacity Benefit, Jobseeker's Allowance, Carer's Allowance, Income Support, Severe Disablement Allowance, Employment and Support Allowance, Widow's Benefit, Bereavement Benefit, Disability Living Allowance, Pension Credit, State Pensions and Attendance Allowance) and records from the Single Housing Benefit Extract.

Other benefits

These include benefits from:

- the National Benefits Database:
 - Incapacity Benefit
 - Jobseeker's Allowance
 - Carer's Allowance
 - Income Support
 - Severe Disablement Allowance
 - Employment and Support Allowance
 - Widow's Benefit
 - Bereavement Benefit
 - Disability Living Allowance
 - Pension Credit
 - State Pensions
 - Attendance Allowance
- records from the Single Housing Benefit Extract

Around 17 million additional records are added to the ABPE by these benefits datasets, the majority of which are from Child Benefit data. As expected, this brings the numbers for children close to the census estimates.

The working age adult population is slightly increased by records with claims on other benefits data, not appearing on PAYE or Tax Credits. The effect is larger for pensioners, where the numbers included are very close to the census estimates for both males and females. This is because of the inclusion of State Pension records on the National Benefits Database, which all but small numbers of pensioners are expected to be eligible for and claim.

The main deficit compared with census now appears to be from age 18 years to the early twenties, where students who are not working will currently be missing.

Patient Register (PR) and Personal Demographic Service (PDS)

The Patient Register (PR) does not show recent activity for most of the population as it is primarily a source for registration. If initial registration occurred in the 12 months prior to the ABPE reference date, this is identified as evidence for activity. New registrations will include large numbers of newborns and new migrants.

Births registrations data is used at a later stage so PR records for those aged zero are not added at this point. For all other ages, those newly registered on PR with no activity already on the BIDS data are added to the ABPE.

For 2016 onwards, Personal Demographic Service (PDS) data is available. This is similar to the PR but updated by more NHS services as shown in [our previous methodology report](#). The PDS data also contain dates of assignment to a GP, which are updated with address changes. Therefore, the PDS can capture more up-to-date activity. This is likely to improve the estimates in areas with high population churn.

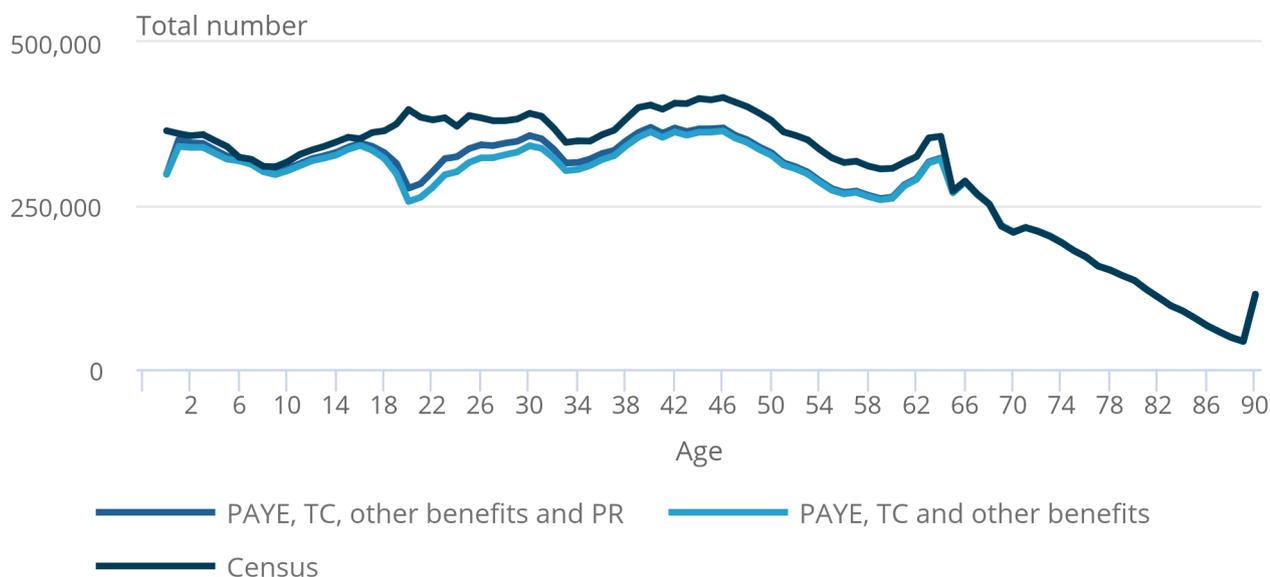
The PDS is used where possible instead of the PR, so the 2016 ABPE V3.0 is constructed using PDS activity whereas the PR is used for 2011. Analysis has shown that the ABPE V3.0 population is about 1.2% lower when the PR is used instead of the PDS for 2016, but this difference may vary substantially between different local authorities and different age and sex groups.

Figure 3a: ABPE V3.0 population after adding active records from the NHS Patient Register, compared with census estimates

England and Wales, 2011, males

Figure 3a: ABPE V3.0 population after adding active records from the NHS Patient Register, compared with census estimates

England and Wales, 2011, males



Source: Office for National Statistics

Notes:

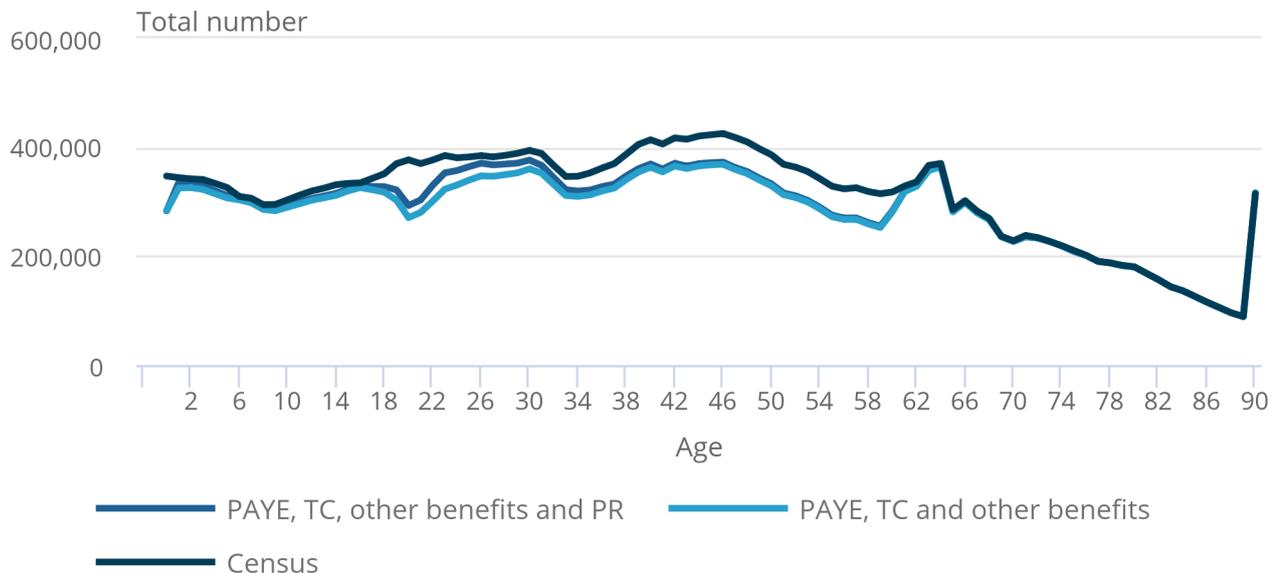
1. ABPE – Admin-based population estimates.
2. "Activity" is defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.
3. PAYE – Pay As You Earn.
4. TC – Tax Credits.
5. Other benefits – these include benefits from the National Benefits Database (Incapacity Benefit, Jobseeker's Allowance, Carer's Allowance, Income Support, Severe Disablement Allowance, Employment and Support Allowance, Widow's Benefit, Bereavement Benefit, Disability Living Allowance, Pension Credit, State Pensions and Attendance Allowance) and records from the Single Housing Benefit Extract.
6. PR – Patient Register.

Figure 3b: ABPE V3.0 population after adding active records from the NHS Patient Register, compared with census estimates

England and Wales, 2011, females

Figure 3b: ABPE V3.0 population after adding active records from the NHS Patient Register, compared with census estimates

England and Wales, 2011, females



Source: Office for National Statistics

Notes:

1. ABPE – Admin-based population estimates.
2. "Activity" is defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.
3. PAYE – Pay As You Earn.
4. TC – Tax Credits.
5. Other benefits – these include benefits from the National Benefits Database (Incapacity Benefit, Jobseeker's Allowance, Carer's Allowance, Income Support, Severe Disablement Allowance, Employment and Support Allowance, Widow's Benefit, Bereavement Benefit, Disability Living Allowance, Pension Credit, State Pensions and Attendance Allowance) and records from the Single Housing Benefit Extract.
6. PR – Patient Register.

Figures 3a and 3b show that, for 2011, the recent PR registrations result in records being added particularly for those aged in their twenties and early thirties, the age groups where international migration is highest. This appears to have a greater effect for females, and numbers of females aged 26 years are only 3.6% below the census estimate, compared with over 10% for males of the same age.

Higher Education Statistics Agency (HESA) data

HESA data provide information about the student population; many of these records may not show activity on any of the BIDS datasets because of students being in education and not employment.

Since registration on the HESA dataset takes place annually for each year of study, presence on the HESA dataset is sufficient to indicate activity in the 12 months prior to the reference date. Records identified as dormant, studying abroad for part of the degree, non-UK students or on a course of less than one year are not included as part of this rule.

English and Welsh School Censuses

Records present on the School Census data are also taken as sufficient evidence of activity since pupils are registered annually. No further filtering of records is possible. This will add further records for children of school age who are not on the Child Benefit dataset.

Only state school pupils are included on the School Census, therefore private and home-schooled pupils will be excluded if Child Benefit is not claimed for them.

In 2013, the [High Income Child Benefit Tax Charge](#) was introduced, meaning effectively no Child Benefit is paid to parents earning above £60,000 per year. As a result, families with a higher income are less likely to apply and information from other data sources for children in these families will be required to ensure coverage of this group.

Birth registrations data

The Births registrations data provide an extra source of activity for children aged under one year at the time of the estimates, since the registration will have taken place within the 12 months before.

For pre-school ages, the main data source currently used is Child Benefit, which will have some undercoverage. The Births data can make up this shortfall for children age zero born in England or Wales. All zero year olds not already appearing on the ABPE from other sources are added.

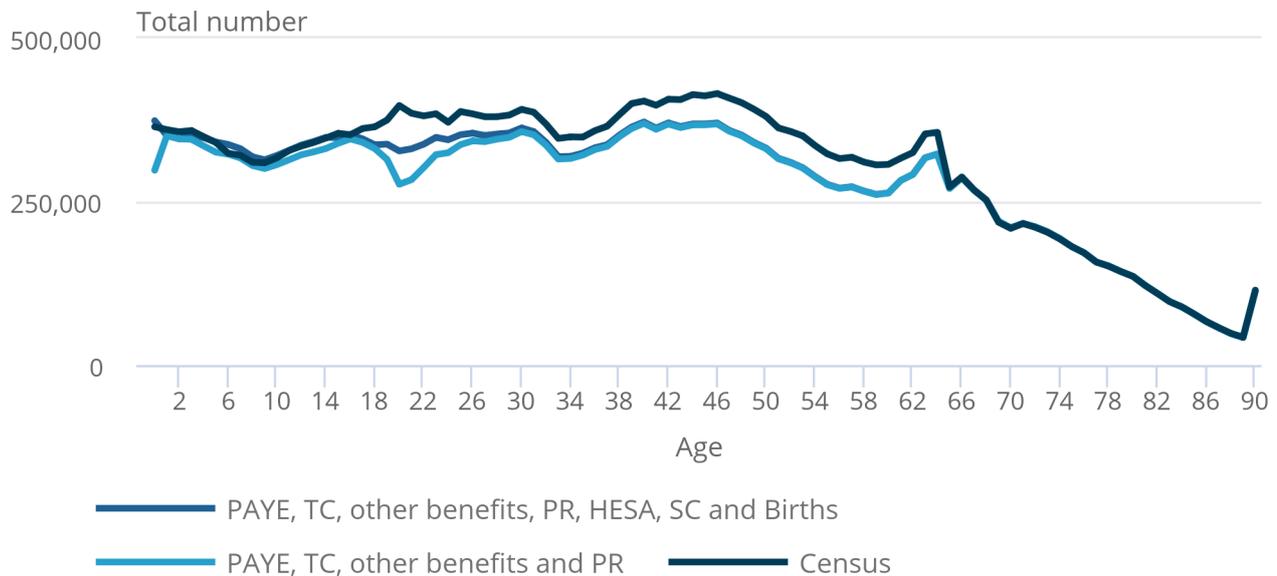
Death registrations data are also available, but are not currently used in the ABPE V3.0 method because of the use of death information on the Department for Work and Pensions (DWP) Customer Information System. We will include death registrations data in the future, this will help remove records resulting from infant mortality ([approximately 2,500 per year](#)).

Figure 4a: ABPE V3.0 population after adding active records from Births, School Census and HESA datasets

England and Wales, 2011, males

Figure 4a: ABPE V3.0 population after adding active records from Births, School Census and HESA datasets

England and Wales, 2011, males



Source: Office for National Statistics

Notes:

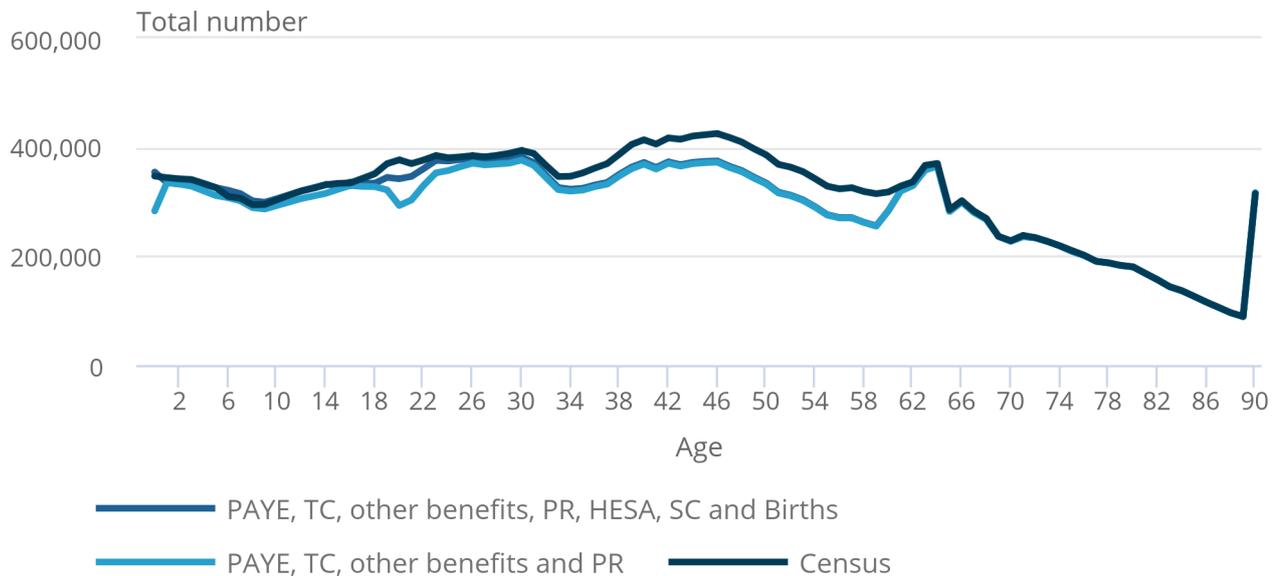
1. ABPE – Admin-based population estimates.
2. "Activity" is defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.
3. PAYE – Pay As You Earn.
4. TC – Tax Credits.
5. Other benefits – these include benefits from the National Benefits Database (Incapacity Benefit, Jobseeker's Allowance, Carer's Allowance, Income Support, Severe Disablement Allowance, Employment and Support Allowance, Widow's Benefit, Bereavement Benefit, Disability Living Allowance, Pension Credit, State Pensions and Attendance Allowance) and records from the Single Housing Benefit Extract.
6. PR – Patient Register.
7. HESA – Higher Education Statistics Agency.
8. SC – School Census.

Figure 4b: ABPE V3.0 population after adding active records from Births, School Census and HESA datasets

England and Wales, 2011, females

Figure 4b: ABPE V3.0 population after adding active records from Births, School Census and HESA datasets

England and Wales, 2011, females



Source: Office for National Statistics

Notes:

1. ABPE – Admin-based population estimates.
2. "Activity" is defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.
3. PAYE – Pay As You Earn.
4. TC – Tax Credits.
5. Other benefits – these include benefits from the National Benefits Database (Incapacity Benefit, Jobseeker's Allowance, Carer's Allowance, Income Support, Severe Disablement Allowance, Employment and Support Allowance, Widow's Benefit, Bereavement Benefit, Disability Living Allowance, Pension Credit, State Pensions and Attendance Allowance) and records from the Single Housing Benefit Extract.
6. PR – Patient Register.
7. HESA – Higher Education Statistics Agency.
8. SC – School Census.

After the addition of these records, the numbers of children included in the new ABPE have become much closer to the census estimates and in some cases are now higher. The deficit for student age groups has been substantially reduced, but there still appear to be some records missing at these ages.

The main deficit of records is now for the working ages from around age 40 years upwards, where male and female estimates are both around 12% below the census estimates. There is still a substantial difference between males and females for those aged in their twenties and early thirties, where female estimates fall within 1% of the census estimates at age 26 years, but male numbers are over 7% below census.

Inclusion of inactive relatives

Although the ABPE V3.0 method is heavily based on activity data, it is recognised that there are also genuine cases of people without activity. Admin data are usually collected at the individual level and this is true for the BIDS data, so it does not account for people living with others and sharing the income of the whole household. For example, couples where one partner works and earns an income but the other does not.

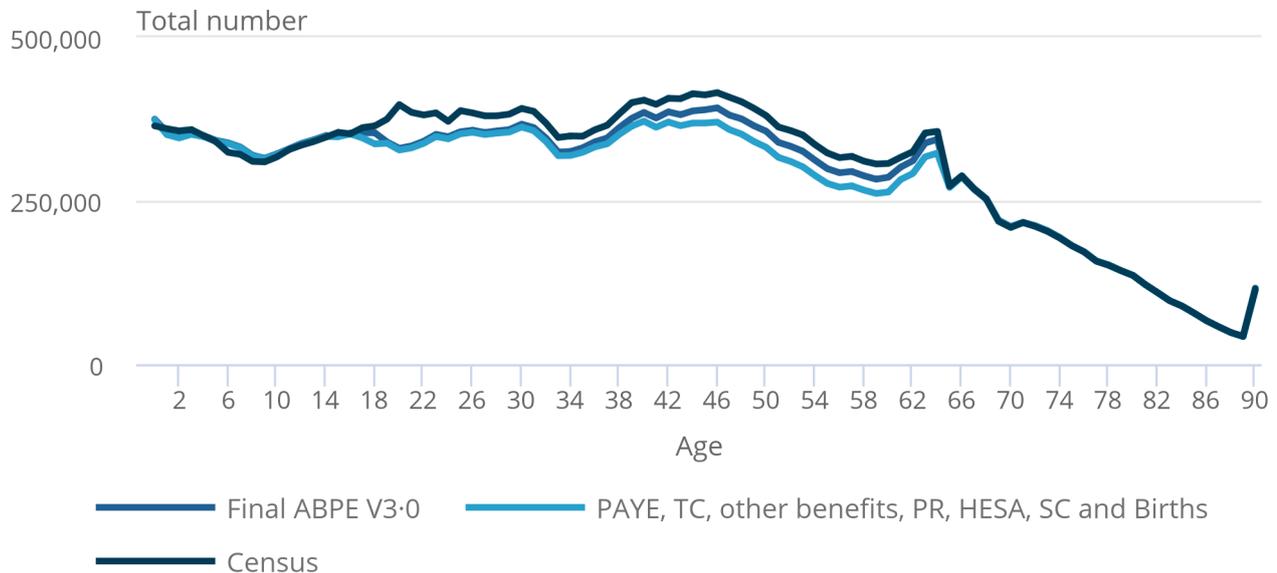
A simple method has been developed to capture many of these inactive partners for couples who are married. This links a record with no activity to an active record in the same address with the same surname. This chooses the oldest active record in the address to prioritise the linking. Here some children are also captured by linkage to parents and the other way around. Rules have been added to exclude adult children currently, since there is more uncertainty around whether they are still present.

Figure 5a: Final ABPE V3.0 population, after inactive relatives are added

England and Wales, 2011, males

Figure 5a: Final ABPE V3.0 population, after inactive relatives are added

England and Wales, 2011, males



Source: Office for National Statistics

Notes:

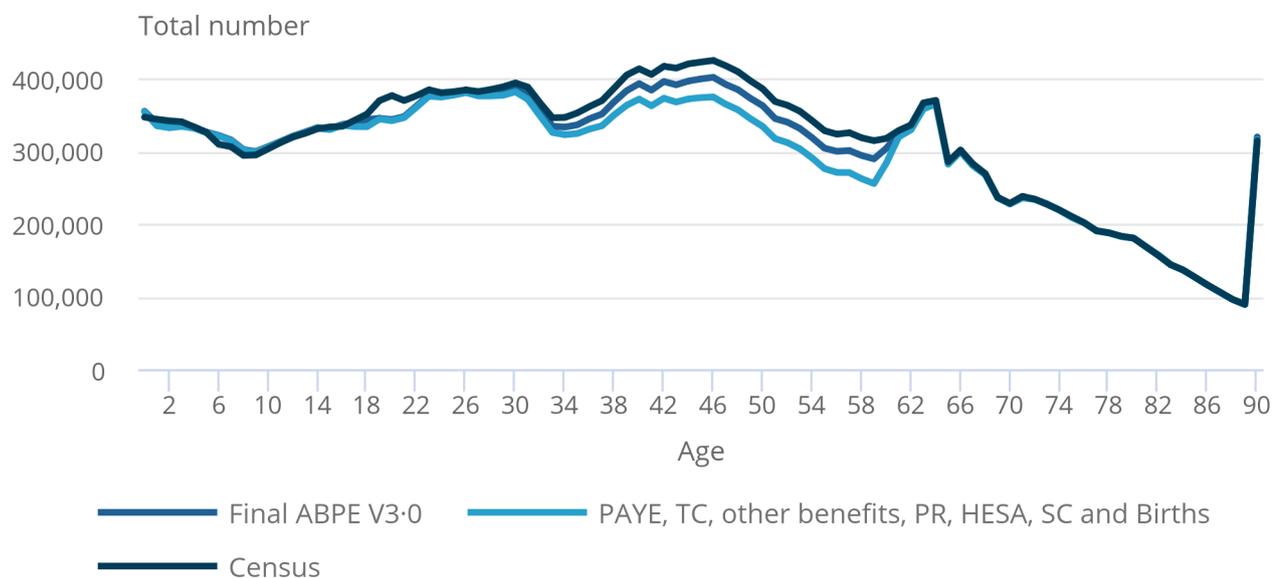
1. ABPE – Admin-based population estimates.
2. "Activity" is defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.
3. Inactive relatives are records showing no activity on admin sources, but share an address with an active person with the same surname.
4. PAYE – Pay As You Earn.
5. TC – Tax Credits
6. Other benefits – these include benefits from the National Benefits Database (Incapacity Benefit, Jobseeker's Allowance, Carer's Allowance, Income Support, Severe Disablement Allowance, Employment and Support Allowance, Widow's Benefit, Bereavement Benefit, Disability Living Allowance, Pension Credit, State Pensions and Attendance Allowance) and records from the Single Housing Benefit Extract.
7. PR – Patient Register.
8. HESA – Higher Education Statistics Agency.
9. SC – School Census.

Figure 5b: Final ABPE V3.0 population, after inactive relatives are added

England and Wales, 2011, females

Figure 5b: Final ABPE V3.0 population, after inactive relatives are added

England and Wales, 2011, females



Source: Office for National Statistics

Notes:

1. ABPE – Admin-based population estimates.
2. "Activity" is defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.
3. Inactive relatives are records showing no activity on admin sources, but share an address with an active person with the same surname.
4. PAYE – Pay As You Earn.
5. TC – Tax Credits.
6. Other benefits – these include benefits from the National Benefits Database (Incapacity Benefit, Jobseeker's Allowance, Carer's Allowance, Income Support, Severe Disablement Allowance, Employment and Support Allowance, Widow's Benefit, Bereavement Benefit, Disability Living Allowance, Pension Credit, State Pensions and Attendance Allowance) and records from the Single Housing Benefit Extract.
7. PR – Patient Register.
8. HESA – Higher Education Statistics Agency.
9. SC – School Census.

Figures 5a and 5b show that the main effect of this method is to add further records where the largest deficit compared with census was seen previously.

It is expected that there are higher numbers of inactive spouses in these age groups. Spouses who are self-employed and hence do not show activity on the BIDS data sources in ABPE V3.0 may also be added by this rule.

The deficit compared with the census estimates has been reduced to around half of its previous size. This suggests that this step is necessary to obtain accurate estimates, but needs further improvement, since spouses not sharing a surname and cohabitating partners will not currently be captured. This may be possible in future using Electoral Register data.

8 . How does the new method compare with official estimates?

Comparison with unadjusted census counts

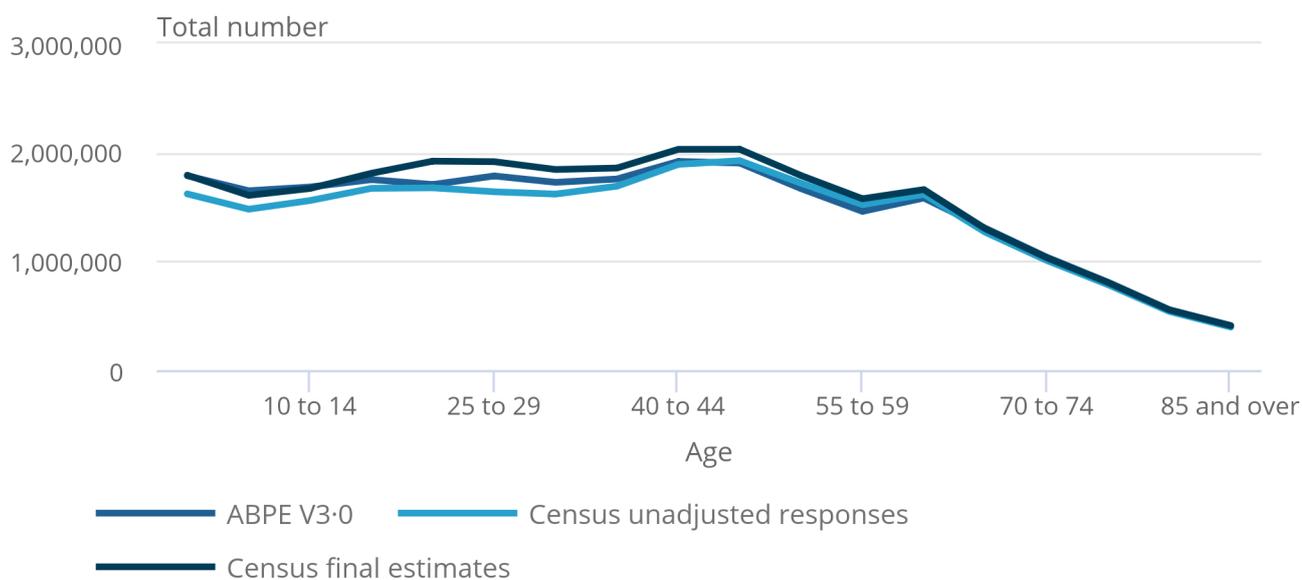
Figures 6a and 6b show how the admin-based population estimates V3.0 (ABPE V3.0) population compares with both unadjusted counts of census responses and the final census estimates, for five-year age groups by sex.

Figure 6a: ABPE V3.0 population, unadjusted census counts and final census estimates

England and Wales, 2011, males

Figure 6a: ABPE V3.0 population, unadjusted census counts and final census estimates

England and Wales, 2011, males



Source: Office for National Statistics

Notes:

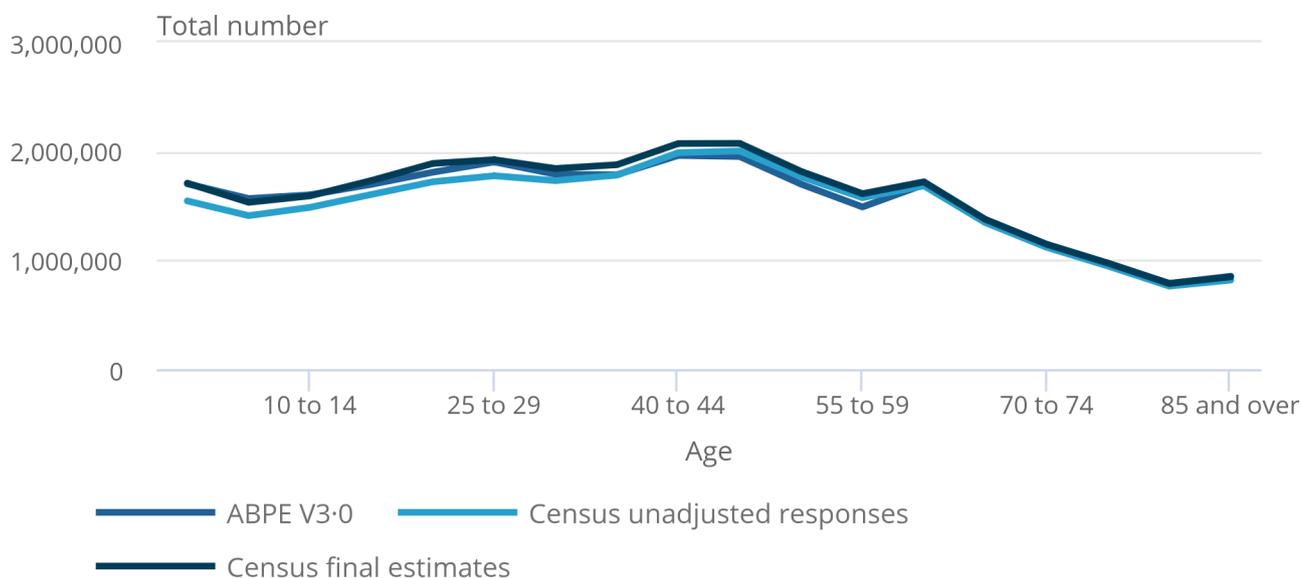
1. ABPE – Admin-based population estimates.

Figure 6b: ABPE V3.0 population, unadjusted census counts and final census estimates

England and Wales, 2011, females

Figure 6b: ABPE V3.0 population, unadjusted census counts and final census estimates

England and Wales, 2011, females



Source: Office for National Statistics

Notes:

1. ABPE – admin-based population estimates.

We have compared ABPE V3.0 with the unadjusted counts of census responses as this is a more like-for-like comparison. An unadjusted census is the number enumerated prior to estimation for non-response, which provides the final [census estimate](#). Our intention is that a Population Coverage Survey (PCS) will measure and adjust for coverage errors seen in the ABPE, much like [the Census Coverage Survey](#) measures and adjusts for census non-response before producing the final census estimates.

For most age groups, the ABPE V3.0 population is higher than the unadjusted census counts, meaning that smaller adjustments by an estimation method would be required than in the census. Around age 40 years, the ABPE V3.0 population dips below the census counts and remains lower for ages up to pension age, meaning larger adjustments would be required for these groups. Above age 65 years, the census adjustments are relatively small and the ABPE V3.0 population closely matches the final census estimates.

For children, ABPE V3.0 is closer to the census estimates than to the census counts. For ages 5 to 9 years and 10 to 14 years, the ABPE V3.0 population is higher than the census estimates, due possibly to some erroneous records still included in the ABPE. Further research is required to determine if some of the records are no longer active.

For example, around 300,000 children are registered on School Census data without recent activity on any other source. Some of these may have subsequently left the country without leaving any indication of this in any data source and are therefore no longer part of the usually resident population.

Therefore, except for ages from 5 to 14 years, it appears that ABPE V3.0 may behave similarly to the unadjusted census counts when combined with a coverage survey and estimation method. This indicates good potential for producing robust population estimates.

Comparison of ABPE V3.0 for 2011 with final census estimates and ABPE V2.0

This section compares the new method with the previous method showing the difference between the approaches. These are compared against the final census estimates to illustrate areas of under and overcoverage.

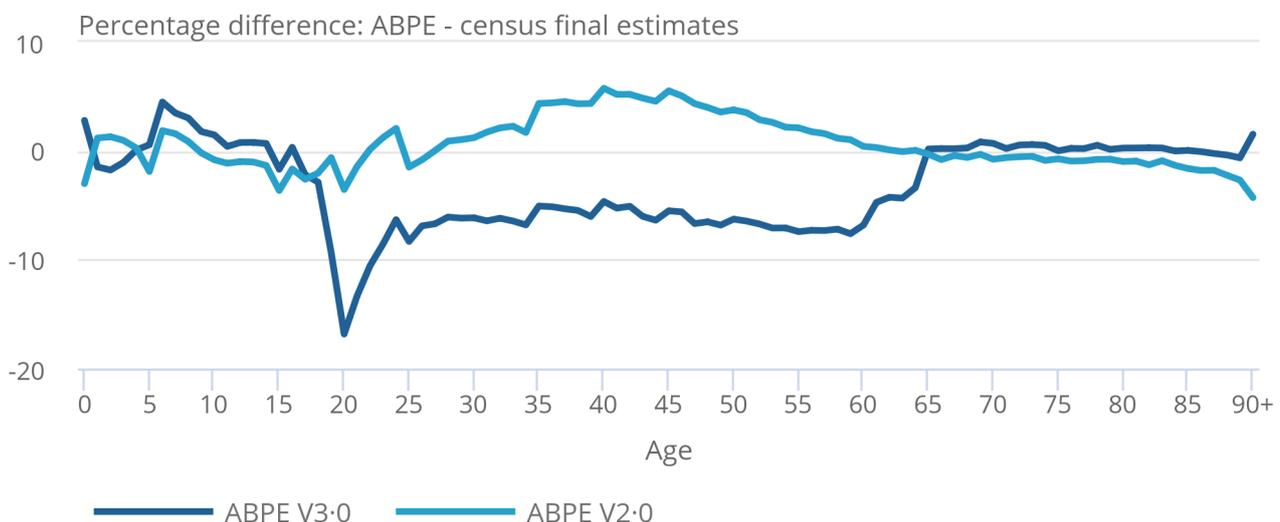
Overall, the population for England and Wales in ABPE V3.0 is around 1.85 million fewer than the 2011 Census and about 1.77 million fewer than ABPE V2, because of the new inclusion criteria. In percentage terms, the ABPE V3.0 population is about 3.3% lower than census, whereas ABPE V2.0 is 0.14% below census. Although this may suggest higher quality for the ABPE V2.0, there are substantial areas of under and overcoverage for different population groups that effectively cancel each other out.

Figure 7a: Comparison of ABPE V3.0 and ABPE V2.0 with final census estimates

England and Wales, 2011, males

Figure 7a: Comparison of ABPE V3.0 and ABPE V2.0 with final census estimates

England and Wales, 2011, males



Source: Office for National Statistics

Notes:

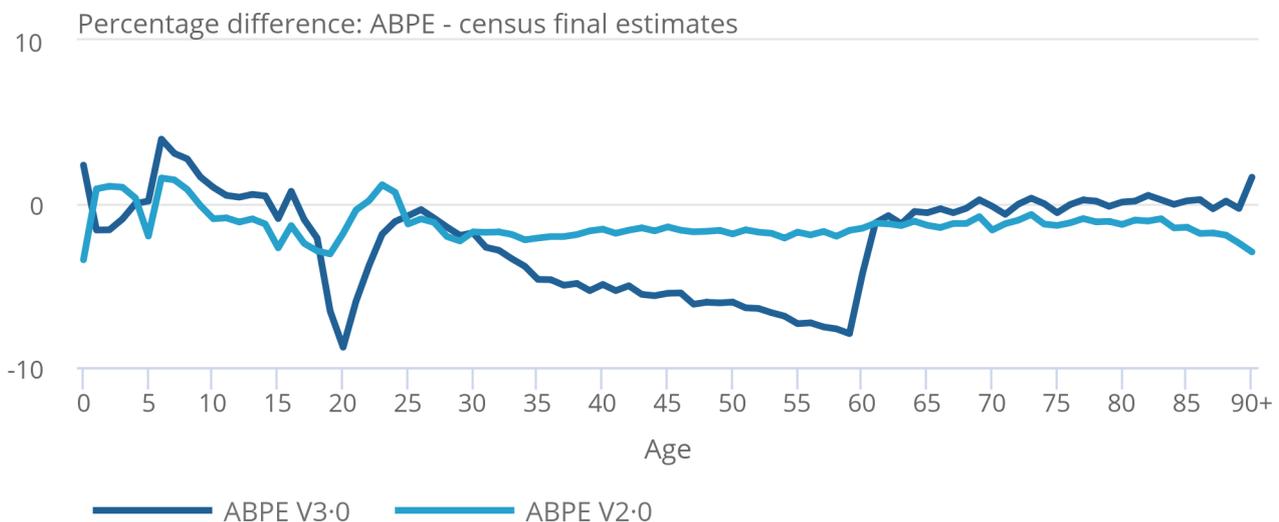
1. ABPE – Admin-based population estimates.

Figure 7b: Comparison of ABPE V3.0 and ABPE V2.0 with final census estimates

England and Wales, 2011, females

Figure 7b: Comparison of ABPE V3.0 and ABPE V2.0 with final census estimates

England and Wales, 2011, females



Source: Office for National Statistics

Notes:

1. ABPE – Admin-based population estimates.

The main flaw in our previous method was overcoverage of working age males, this is no longer present in ABPE V3.0. The method has been successful in its principal objective. The remainder of this section discusses the differences between the two methods and the census for each specific age group.

For children aged under one year, the new method shows overcoverage compared with the census, this is a contrast to our previous method where we saw undercoverage for this group.

The ABPE V3.0 estimate is derived using births registrations and active Child Benefit records. As with children registered only on the School Census, some of those recorded only on the births data may have since moved abroad, with no sign of this departure appearing in the data. To prevent this, it may also be useful to link the children and parents in the ABPE, so that evidence of the parents emigrating can also be used to exclude their children from the population.

As mentioned previously, infant mortality may also contribute to overcoverage for zero-year olds and the effect of this can be minimised by linking to death registrations in future versions.

For children aged between one and three years, ABPE V3.0 estimates are lower than both census and ABPE V2.0, and this is likely to be because of Child Benefit not being claimed for all. There are no other activity data sources with high coverage currently available for this age group, with only small numbers likely to have a registration on the Patient Register (PR) within the last year.

A future development for this may be to identify further records of linked children and parents using the details recorded in the Births Register. If a parental record shows evidence of activity and is included in the ABPE, the linked child record could also be included.

For children of school age, ABPE V3.0 is higher than ABPE V2.0 and above the census estimates as discussed in the previous section. This peaks at age six years around 4% higher than the census estimates for both males and females.

Although the estimates are higher, the pattern shown is similar to ABPE V2.0. Further research is required to determine if some of these additional records are no longer present, for example, by looking for evidence of de-registration from NHS services.

At ages 19 to 23 years, we see a sharp dip below the census estimates for ABPE V3.0, which does not occur for ABPE V2.0. This age range is generally a period of transition from education to employment, perhaps also including periods of unemployment or economic inactivity. This makes it a complex period to capture accurate activity data, as this relies on individuals updating their records. We are exploring other potential data sources that may help capture activity over this period, such as further education or apprenticeships data, which are not currently available for use in the ABPE.

Benefits may be claimed for periods of unemployment, but the initial rules for the benefits datasets will currently exclude claims that end shortly before, or start shortly after, the reference date of the ABPE. Others may have periods of inactivity, but the method for adding inactive relatives currently excludes adult children because of high uncertainty around such records and the potential for overcoverage. Therefore, additional data sources, and further research and development of these initial rules will bring improvements for ages between 19 and 23 years.

For ages from 19 to 34 years, there are significant differences between male and female estimates. Initial analysis of the Benefits and Income Data (BIDS) sources revealed that the majority of Tax Credits and Housing Benefits claimants are female and the differences are greatest in this age range. Therefore, this method is better at capturing females at this age than males with the data currently available. [Census statistics on self-employment](#) show that males comprise a majority of the self-employed, so an important next step is to include self-assessment data to try and reduce the difference we see.

Similar coverage of males and females is seen for ages from the mid-thirties up to pension age, with estimates between approximately 5% and 8% below census estimates.

For pensioners, relatively accurate estimates were obtained in ABPE V2.0, but the new method improves on these to give estimates very similar to the census.

Comparison of ABPE V3.0 2016 with official mid-year estimates and ABPE V2.0

For 2016, we can similarly compare the patterns of coverage in both ABPE methods and the [official mid-year population estimates](#).

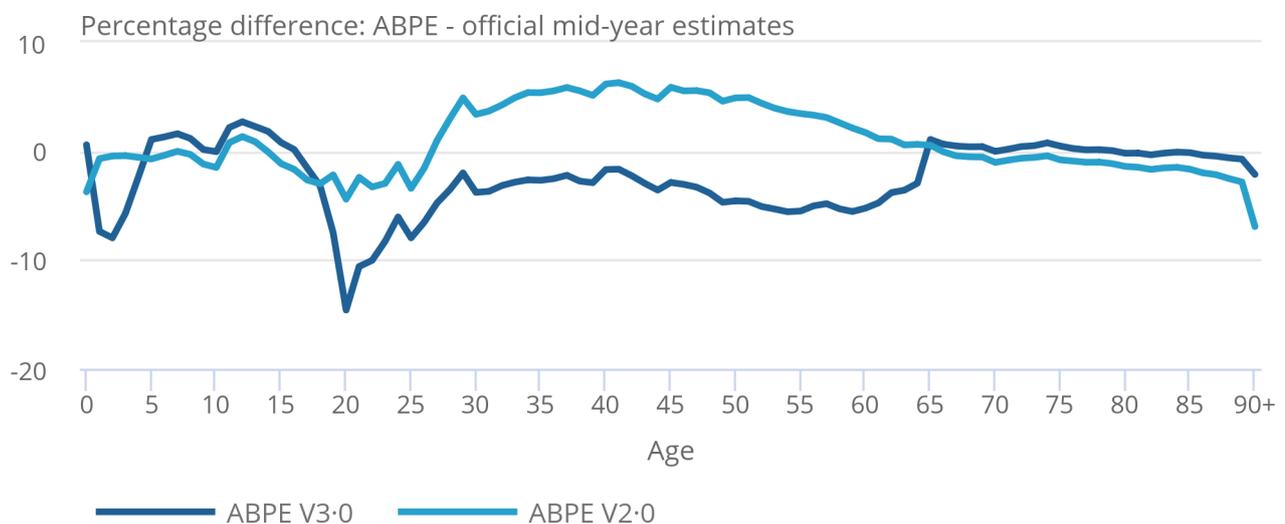
For 2016, the new method has a population around 1.13 million lower than the official estimates, a difference of around 1.94%. In contrast, the ABPE V2.0 population is higher than the official estimates, by 384,000 people or around 0.66%.

Figure 8a: Comparison of ABPE V3.0 and ABPE V2.0 with official mid-year estimates

England and Wales, 2016, males

Figure 8a: Comparison of ABPE V3.0 and ABPE V2.0 with official mid-year estimates

England and Wales, 2016, males



Source: Office for National Statistics

Notes:

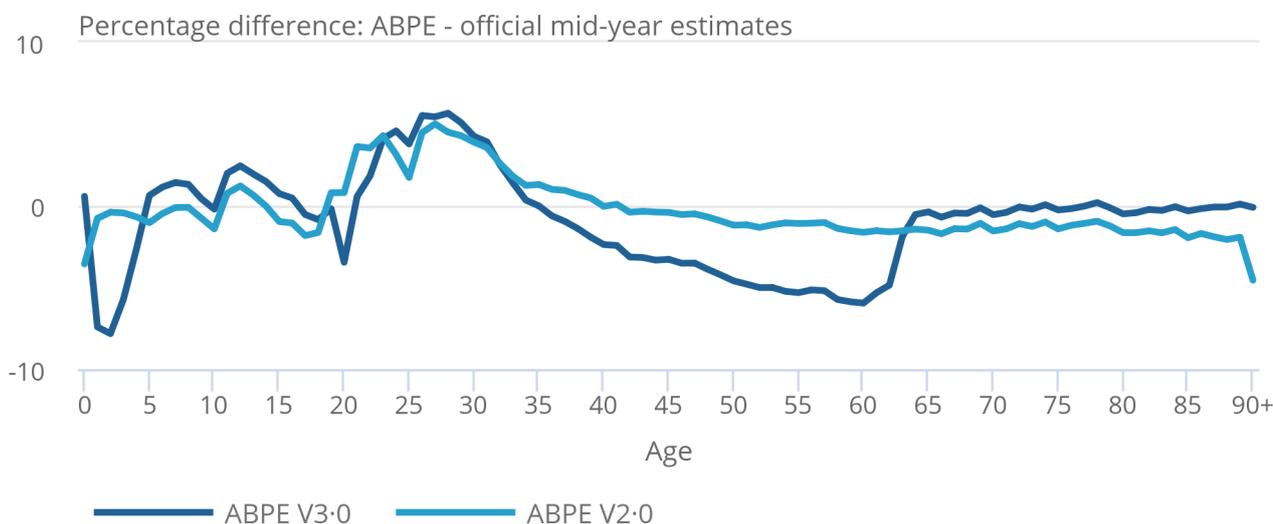
1. ABPE – Admin-based population estimates.

Figure 8b: Comparison of ABPE V3.0 and ABPE V2.0 with official mid-year estimates

England and Wales, 2016, females

Figure 8b: Comparison of ABPE V3.0 and ABPE V2.0 with official mid-year estimates

England and Wales, 2016, females



Source: Office for National Statistics

Notes:

1. ABPE – Admin-based population estimates.

The main change from the 2011 estimates is for females, with the new estimates now higher than the official mid-year estimates for ages from 21 to 34 years, reaching over 5% higher at ages 26 to 28 years. The difference between males and females in this age range is even larger than for ABPE V3.0 in 2011.

Using the Personal Demographic Service (PDS) instead of the Patient Register (PR) may account for some of this difference. As seen in Figure 3, the contribution from PR records is largest in this age range and is larger for females than males. The higher activity shown by the PDS may accentuate this difference further.

The ABPE V3.0 estimates for females aged 21 to 34 years also show a high similarity to those of ABPE V2.0. It is possible that the ABPEs contain some erroneous records because of individuals who have emigrated or are short-term international migrants intending to stay for less than one year and therefore who are not included in the official estimates. However, the mid-year estimates suffer gradual drift in the years between censuses.

While the effects are mainly expected to be seen in estimates for subnational areas rather than giving large differences at the national level, this may offer some explanation to the difference seen in both ABPEs. It is therefore difficult to account for these differences without much more detailed research.

ABPE V3.0 estimates for school age children are again higher than the official estimates, but to a lesser extreme than in 2011. Also similar to 2011, the ABPE V3.0 population is higher than ABPE V2.0 for this age group. Single source records from Child Benefit and School Census again contribute additional records, but for 2016 we also use the NHS Personal Demographic Service (PDS) instead of the Patient Register (PR), which shows more activity¹.

For children aged one to three years, there is a much larger difference between the ABPE V3.0 and official estimates than seen for 2011. This may be caused by Child Benefit being claimed for fewer children because of the [High Income Child Benefit Tax Charge](#) introduced in 2013.

Notes for: How does the new method compare with official estimates?

1. Provisional ABPE V3.0 estimates for ages 5 to 15 years are shown in our previous [slide share](#). These were constructed using only School Census, Child Benefit and Patient Register (PR) as activity sources. The final ABPE V3.0 estimates shown in this report are higher because of using the NHS Personal Demographic Service (PDS) instead of the PR, and also because further children without activity are added through sharing an address with an active parent with the same surname.

9 . Summary and next steps

This is the first attempt to produce admin-based population estimates (ABPE) using activity data and additional rules to determine inclusion in the population. This has been largely successful in its main goal of reducing the overcoverage seen in previous research. Although the new method produces a population that is lower than the official estimates for most age groups, there are still expected to be some incorrect records included. Bringing in additional data sources in the future will help with this.

The similarity between the coverage of admin-based population estimate V3.0 (ABPE V3.0) and the unadjusted census responses indicates that the new method should be suitable for producing robust coverage-adjusted estimates using a Population Coverage Survey (PCS). This is using a similar method used to successfully adjust the census.

The inclusion of further data sources may allow indicators of activity to be found for records that are not currently included, such as those with only self-employment activity. We also continue to work closely with data suppliers to understand administrative data sources and how we can build these into our future systems, filling existing gaps and providing stronger evidence to include or exclude records from the ABPEs.

The wider use of available relationships information is also likely to improve the ABPE, such as linking children to parents. This would allow additional children who are not on Child Benefit or at a state school to be included. It would also give more consistency in the ABPE, by ensuring that children are not included in a household without parents or guardians, potentially reducing the number of incorrect records.

Although the current focus is on designing activity-based rules to determine which records to include in the usually resident population, an important aspect is to ensure the correct records are allocated to the correct area to provide robust population estimates at the subnational level. Therefore, we will also continue to develop this method to provide robust estimates at the local authority level.

We are also progressing work to provide uncertainty intervals for the ABPEs, offering an independent quality check for these estimates.

This research provides a platform for combining with a PCS to produce coverage-adjusted estimates in a similar way to the method used to successfully adjust the census. We are undertaking work to develop and test this use of a PCS in order to evaluate the quality of the ABPE. This will also be part of producing coverage-adjusted estimates in conjunction with an estimation method. Further research will also work towards creating a framework to allow evaluation of error throughout the process of constructing an ABPE and to create uncertainty intervals for the estimates.

10 . More about our transformation journey

- More information is available about the [transformation of the population and migration statistics transformation system](#).
- [Research Outputs](#) on using administrative data to produce population statistics.
- Previous research using administrative data to produce [estimates on the size of the population](#) (previously Statistical Population Datasets (SPDs)).
- August 2017 report on our progress towards developing a better understanding on [student migration to and from the UK](#).
- May 2018 [update of the Migration Statistics Transformation Programme](#).
- A [report on international migration data sources](#) setting out how we are using Home Office administrative data to further our understanding of international migration.
- [February 2019 report: Understanding different migration data sources, a workplan](#). Examining the issues with comparing the UK's various migration data sources and our plans to explain the differences between these sources.
- January 2019 report [updating our users on our population and migration statistics transformation journey](#) and seeking feedback on our future plans.

11 . More about our official population and migration statistics

- [Population estimates for the UK, England and Wales, Scotland and Northern Ireland statistical bulletins](#).
- All information and publications on [international migration](#) produced by the Office for National Statistics.
- All information and publications on the [size of the population](#) produced by the Office for National Statistics.