The assessment of the Office for National Statistics' Admin Based Population Estimates - Independent expertise

Arkadiusz Wiśniowski, The University of Manchester

Email: a.wisniowski@manchester.ac.uk

2 July 2024

Summary of recommendations

Recommendations I consider essential for the ONS to address:

R1. To provide a comprehensive and detailed methods guide that will ensure that the Dynamic Population Model (DPM) is reproducible. The guide should describe in detail:
   a. data inputs,
   b. modelling framework,
   c. assumptions regarding population components,
   d. computational methods,
   e. model testing, and
   f. analysis of the outputs.

   The methods guide should contain versioning similar to the versioning of the Statistical Population Dataset (SPD).

R2. To provide in the documentation (R1) a clear differentiation between bias and accuracy (or precision) of the data inputs and assess each data input in terms of bias and accuracy. The assessment should inform the DPM. Such a distinction is essential for the DPM to produce reliable (i.e. unbiased and accurate) population estimates.

R3. To quantify in the documentation (R1) the assumptions in the model, e.g., for precision this could be done by providing coefficients of variation around the mean, rather than stating that one source is more precise than the other. The current version of the DPM relies on informative priors and such quantification is required as an input to the model. It will ensure that the various assumptions can be tested and their impact on ABPEs assessed.

R4. To test and document the impact of using a coverage benchmark in the DPM (Option 1: correct in the data inputs, Option 2: Correct in the DPM via model parameters). The documentation should contain a description which option has been implemented.

R5. To analyse the sensitivity of the ABPEs to a variety of prior distributions assumed for the accuracy (precision) of each of the data inputs. Special attention should be paid to precision of migration (currently internal, cross-border and international migration being jointly modelled as in- and out-flows to and from LAs). Sensitivity analysis should be carried out for the prior distributions for the coverage adjustment parameters. These analyses will inform if

the ABPEs are robust to the assumptions about data quality and help identify extreme situations where the DPM may require further research.

R6. To continue developing a quality assurance processes at each stage of producing ABPEs, i.e. starting with producing data inputs, assessment of their bias and accuracy, quantification in terms of data-corrections and/or model parameters, as well as robustness and sensitivity analyses of the DPM and ABPEs. This is to ensure the sustainability of the DPM if data inputs change or new sources are introduced in the future.

R7. To provide a statement that accompanies the DPM-based ABPEs on the potential sources of uncertainty or bias that are unaccounted for and, where possible, an assessment of their importance in a given situation, e.g. when considering estimates for age groups or LAs.

Further recommendations:

R8. To continue research on the data source(s) to be used as a coverage benchmark for the admin-based data used in the DPM (to inform R4).

R9. To develop a process of assessing the quality of all data inputs in terms of bias and accuracy of the data inputs for the purposes of being used in the DPM. Such a process could provide a structure for the data quality assessment (as described in R2) if data collection mechanisms change or new data sources are introduced.

R10. To continue research to better understand the nature of errors (biases and uncertainty) in the data sources that are used as inputs to the DPM. This includes continuing and documenting the simulation studies and other assessments of the data inputs, e.g., the results of simulations carried out for the inclusion rules in the Demographic Index and SPD. This should be done in consultation with a variety of stakeholders and will inform assessments in R2, R3 and R9.

R11. To continue testing of the DPM and resulting ABPEs by using goodness of fit measures via prior and posterior predictive checks, which are a typical component of a Bayesian workflow. These checks include predicting the data inputs by using a model with only prior distributions, or estimated by using data inputs, potentially perturbated by random or systematic removals of portions of the data. Such analyses would complement the sensitivity tests as described in R5.

R12. To develop a comprehensive battery of tests (based on R5 and R11) that can be automatically applied to future versions of the DPM, e.g., if hierarchical structure is to be included in it.

R13. To continue developing an interactive R package containing a toy model that would demonstrate the workflow of the DPM and permit testing of (some of) the model assumptions. This would also benefit the communications of the estimates especially to stakeholders interested in a more detailed understanding of the model. It will also permit a more informed consultation with stakeholders if changes in the model or model assumptions are to be introduced. Feedback from stakeholders may also lead to revisions of the DPM.

R14. To develop an interactive dashboard[1] (that would accompany the R package or be a standalone piece of software) that would enable comparisons of the ABPEs estimates with

[1] An example of such dashboard that enables a (visual) comparison of model results is https://maciej-jan-danko.shinyapps.io/HMigD_Shiny_App_I/.

the 2021-Census-MYE and also for various assumptions in the DPM. The visualisations could be accompanied by the estimates of errors (bias, e.g., via Mean Percentage Error, and precision, via e.g. Mean Absolute Percentage Error) and a glossary and links to documentation.

R15. To engage with the local authorities and government stakeholders to learn about specific characteristics of the areas and types of errors that may be present in the admin data regarding these areas. This could inform the DPM development by, e.g., helping create the typologies of areas that could then be included in the DPM.

R16. To provide a justification for the current implementation of the DPM that relies on a simplified demographic assumptions because of the computational difficulties. This could be done by providing a comparison of the full model implemented on a small scale with the simplified approach followed by an analysis of errors, similarly to how the assessment of the computational methods and their impacts on the ABPEs has been done.

For consideration:

R17. To continue research into methods for quantifying uncertainty in the admin-based migration statistics (international and internal) as these are crucial to producing reliable ABPEs.

R18. To maintain or develop within the ONS the capacity for implementing computational methods used in and around the DPM to reduce risks related to future changes in software packages used in preparing current implementation of the DPM. This will ensure the sustainability of the DPM.

R19. To consider using other countries mirror migration statistics, especially for the British nationals where the International Passenger Survey is being used due to the difficulties of capturing them in the admin data.

R20. To consider a risk assessment of under- vs over-predicting population and its distribution by age, sex and local authority. Such exercise could inform stakeholders about population characteristics of concern and may also guide the research into data quality and representation of uncertainty in the DPM.

Background

1. The purpose of this report is to assess the suitability and quality assurance of the data and methods used in the Dynamic Population Model (DPM) to produce admin-based population estimates (ABPEs) in England and Wales. This report covers aspects of data used as inputs to the model, the modelling framework and its strengths and limitations in producing population statistics, as well as the transparency and sustainability of the methods.
2. The official statistics on population size in England and Wales by age, sex and administrative areas (Local Authorities - LAs) are produced based on the population balancing equation, where the starting point (i.e. a population stock on a given day) is defined by the Census, which is then updated (or rolled forward) to 30 June (or mid-year) for each year, with information on births, deaths and movements into and out of the area; special populations such as members of armed forces and prisoners are also accounted for (e.g., ONS 05/09/2022, 23/11/2023b). In England and Wales, we distinguish movements to/from abroad (international migration), to/from constituent countries (cross-border migration), and between areas within England and Wales (internal migration).

3. These current official statistics are referred to as Mid-Year Estimates (MYE). The common issue with these estimates based on censuses is that they are precise in the census year and their quality deteriorates the further we depart from the census, which is known as an intercensal drift (Blackwell et al. 2022, ONS 05/09/2022), also called an unattributable population change (UPC). This drift or UPC can be caused, in principle, by an error in the population baseline from the preceding census (census base) and by the error in the net migration estimates used in creating mid-year population for 2002-2011 (ONS 12/07/2012)[2]. Similar concerns have been raised about post-2021 Census MYE, especially at a local authority level (Champion 2024).
4. Since the above issues are well-known to the Office for National Statistics, they have been researching, since before Census 2011 (Blackwell et al. 2022) alternative methods for producing population estimates based on administrative data sources. Currently, these new admin-based population estimates (ABPEs) are considered official statistics in development, previously known as "experimental" (e.g., ONS 23/11/2023a). The ABPEs are produced by using a Dynamic Population Model, which is a statistical model that enables estimation of population stocks and flows and accommodates the accounting identity[3] at a national as well as sub-population level, e.g. in a particular area. The model takes as input a variety of administrative data on population stocks and flows. The DPM utilises Bayesian inference to estimate model parameters and produce population estimates with measures of uncertainty – credible intervals (Bayesian counterparts of frequentist confidence intervals).

DPM data inputs

5. In the analysis of the quality of the data, the ONS follows the guidelines of the European Statistical Systems that set the dimensions of data quality: relevance (meeting user needs), timeliness, coherence with other sources and internal comparability, accuracy and reliability, output quality and accessibility. Each of the DPM data inputs are assessed against those criteria (ONS 25/05/2023, 29/12/2023, 23/02/2024). Further, the development of a Quantitative Quality Indicators (QQI) to quantify the quality of the administrative data is exemplary and potentially providing key information about the data quality to the DPM, which is crucial for the outputs, as I explain later. However, I noted that the terms related to accuracy and bias are not always clearly described. For instance, more technical reports (Law et al. 2022) differentiate between bias (a systematic error) and accuracy (variance, or uncertainty, around an estimator) and clear guidelines are created for assessing the quality of population estimates (ONS 2023c), whereas other documents (e.g., ONS 29/12/2023, 23/02/2024) tend to use a broader definition of accuracy that encompasses both bias and uncertainty. I thus recommend a distinction between bias and accuracy (i.e. uncertainty) is provided in the documentation, especially for the purposes of informing the DPM about the quality of the data.
6. The inputs to the DPM are population stocks, such as a baseline population estimated from the 2011 Census for 30 June 2011 (i.e., 2011 Census-based MYE) or snapshots of population counts captured in the administrative sources on a given day, and flows, that is counts of births and deaths (natural change) as well as international, cross-border and internal migration usually between 1 July and 30 June of a given year. The descriptions of the data inputs are provided within the documentation of the model (e.g., Elliott & Blackwell 2023, ONS 14/07/2022) and through the websites documenting the developments of the model

[2] The 2011 Census estimates were around 476,000 higher than the rolled-forward MYEs (ONS 12/07/2012).
[3] The accounting identity is satisfied approximately because of the use of open ended last age group.

(e.g., ONS 27/06/2023b). However, the details of their production are somewhat difficult to navigate because of the extensive cross-referencing to previous documents/releases, and potentially also due to the rebranding of the previous versions of the ABPEs before the deployment of the DPM in July 2022 into Statistical Population Dataset, currently used as an input to the DPM (ONS 27/06/2023a). The ONS confirmed that a methods guide and an interactive R package will be published that will contain each model specification. This should enable stakeholders to better understand the details behind the ABPEs production.

Stocks

7. MYEs based on the 2011 Census are used as a baseline input to the DPM. The MYE that follow census were adjusted for errors in the components of population change in a process called rebasing and reconciling (details available in ONS 30/11/2017). The 2021 MYEs stock estimates derived from the (rebased) 2021 Census have been used as an input to the model, as well as a gold-standard benchmark for assessing the ABPEs based on the DPM, without using them in the DPM (ONS 28/02/2023a). The rebased MYEs 2012 to 2020 were not used in the model. There is a general consensus that the rebased Census estimates are the best population measures available (e.g., ONS 18/12/2023) but 2021 Census may have some limitations due to it having been carried out during the COVID-19 pandemic (ONS 23/11/2023c). Patient Register was also used as a stock in 2012-2015 and Personal Demographics Service (PDS) that is based on the registrations with a GP (ONS 18/12/2023, Elliott & Blackwell 2023).
8. MYEs used as input to DPM are subject to statistical uncertainty, which reflects their accuracy. These are usually expressed in terms of confidence or credible intervals. Census base from the 2011 Census had confidence intervals for population of England and Wales of +/- 0.148 per cent (ONS 2012). This uncertainty is derived from the Census Coverage Survey (CCS). MYEs are created based on this Census base and accumulate uncertainty over time, the further the year from the census, which is referred to as intercensal drift (ONS 25/05/2023, 29/12/2023, 23/02/2024; see also Point 20). Most of the uncertainty in population estimates further from the census year comes from international and internal migration (ONS 27/07/2020).
9. Population stocks used as inputs to the DPM also include Statistical Population Dataset (SPD), referred to in the most recent ONS documentation as Version v4.1 (ONS 18/12/2023b; various previous versions have been used in the preceding publications by the ONS). As mentioned earlier, the SPD was previously referred to as ABPE but was rebranded to reflect its status as the DPM input (ONS 27/06/2023a). The SPD has an ambitious goal of providing the approximate measure of the resident population derived from a variety of administrative data sources. Since the calculation of the SPD would be for each year, the risk of drift observed in the MYE would be lower. The SPD is one of the key inputs to the DPM. The data sources used to create the SPD are well-documented (ONS 03/03/2023). The comparison of the SPD (v4.0) for year 2021 with the 2021 Census estimates showed that while the age-sex and area (LA) profiles are generally similar, there are some considerable differences that will require further research (ONS 28/02/2023a).
10. The “backbone” of the SPD is a Demographic Index (DI), which is a dataset containing linked data on individuals in administrative sources (ONS 2022). The DI records are then included in the SPD if they meet set inclusion criteria based on activity of individuals recorded in the data (through interacting with the administrative systems). However, it is acknowledged that the linkage process for the DI creation and the filtering of the SPD may overlap (ONS 2022), which may introduce error to the data. Furthermore, an exercise of linking the DI records with the Census Coverage Survey (CCS) showed that only a small proportion of the CCS

respondents (less than 1%) were not linked to the DI (ONS 01/03/2023). However, the analysis of the linked data showed that young males, those in London, or those not born in the UK or speaking English as their first language were more likely to be missing from the DI. As acknowledged by the ONS (2023a), more work is needed to improve the linkage methods and better understand the quality of linkage that underlies the DI and its potential impact on biases and uncertainty that can be propagated into the SPD and, subsequently, DPM.

11. The main problem with the SPD is overcoverage (e.g., double counting of individuals in the admin data or the inclusion of individuals who are not usual residents but appear active in the admin data), which seems to be more difficult to handle in estimation processes than undercoverage (i.e., not including some of the usual residents in the data; see Law et al. 2022, 2023). Various strategies and methods have been proposed to improve the quality of the SPD in terms of coverage. The adjustment for over- and undercoverage takes place in the creation of the SPDs (Law et al. 2023) and it also is implemented in the DPM via model parameters (Elliott & Blackwell 2023). This is one of the key aspects of the model, as I discuss later (Point 25).
12. The problem of under- and overcoverage in the SPD was analysed in an exercise of linking the SPD (v4.0) in 2021 to the Census 2021 and Census Coverage Survey (CCS; ONS 28/02/2023c). It was found that 7.3% of those on 2021 Census and CCS were incorrectly excluded (i.e., undercoverage) from the SPD v.4.0, while 8.6% in the SPD were incorrectly included (i.e., overcoverage). While the differences in under- and overcoverage may cancel out at aggregate levels similarly to what has been shown by Champion's (2024) analysis of the UPC, care needs to be taken because these two issues may affect populations with varying characteristics (age, sex) in different areas, as demonstrated on the example of Harrow LA, where differences in incorrect inclusions and exclusions across age groups were found (ONS 28/02/2023c). In another comparison of the SPD v4.0 (ONS 27/06/2023a), it has been found that the difference between coverage-adjusted SPD and the 2021-Census-based MYE is nearly 4% at a national level, and even larger relative differences were found for detailed characteristics. I thus suggest that in the future developments of the DPM, coverage parameters are construed in a way that reflects separate issues related to the under- and overcoverage. This would potentially reduce bias in detailed population characteristics, especially for areas with higher population churn or age profiles of young working age populations. Further, in the current version of the DPM, the model coverage parameters are created by using 2011 and 2021 Census-based MYE and a linear interpolation (ONS 28/02/2023a). I advise that if the SPDs are corrected by using, e.g., 2021 Census data (benchmark) and used as an input to the DPM, then the same benchmark is not used to inform coverage parameters in the DPM. Otherwise there is a risk of the DPM over-correcting for the coverage issues; this will also violate an assumption of not using data twice within Bayesian inference (Gelman 2008, Robert & Ntzoufras 2012). This has been considered by the ONS in their scoping of a data source to be used as a coverage benchmark (Law et al. 2022: Figure 1).

Flows

13. Data on births and deaths in England and Wales are sourced from the Civil Registration System administered by the ONS and are of very high quality compared with migration data, despite minor delays in reporting (ONS 23/02/2024, 29/12/2023). In the DPM, births and deaths are considered error-free and the only uncertainty related to them comes from the fertility and mortality rates through the population at risk in the denominator (Elliott & Blackwell 2023), which is a reasonable assumption.

14. The production of international migration data (long-term international migration, or LTIM) has been going through changes since the COVID-19 pandemic, when the main source of information, International Passenger Survey (IPS), was suspended. The new methods rely, to a much greater extent, on the administrative sources, such as Home Office Border and Immigration Data (HOBID, linked visa and travel data), Registration of Population Interactions Database (RAPID) that builds upon National Insurance Numbers, and statistics from HESA. The IPS is still used to produce estimates of migration of British nationals (ONS 03/05/2024) yet new methods for using administrative data are being developed. Also, a new way of including asylum seekers in the admin-based migration statistics had to be developed.
15. Historically, international migration was subject to high uncertainty, especially for detailed characteristics such as age, sex, country of origin and LA where migrants reside. When relying on the IPS, it also referred to a definition of intended migration (i.e. when a person arriving in the UK or departing from the UK intended to stay in – or outside – of the country for more than 12 months), whereas the new sources permit, in principle, estimating the actual migration. This complies with the UN definition of an international migrant[4]. This, however, causes a delay in providing statistics as persons need to stay in the UK as usual residents for the 12 months before they are recorded as migrants in the database. This limitation is overcome by developing and providing provisional migration estimates, which is a sound strategy that can satisfy stakeholders at the expense of potential corrections to the provisional estimates once the official data arrive. For this purpose, advanced and novel methods are being developed and applied by the ONS (ONS 16/04/2021).
16. As mentioned in the previous paragraphs, the IPS estimates were subject to the sampling error and were produced with measures of uncertainty. The theoretical foundations for the measures of uncertainty of the admin-based migration estimates (ABMEs) are still under development (ONS 01/06/2023) but the hope is that because of the reliance on the admin data, the uncertainty of the estimates can be reduced. However, the currently provided measures demonstrate substantial uncertainty for selected flows. For instance, the 95% uncertainty interval (based on adjustments and modelling) for EU national immigration in 2022 was (112,800; 195,600), whereas for emigration it was (151,000; 270,600) (ONS 01/06/2023; Table 5). The width of the interval suggests that the net migration of EU nationals can be both positive and negative, if we assume there is no correlation between the two flows. Further, the uncertainty of migration of British nationals is relatively lower, despite them relying on the IPS, which would be expected to yield more uncertain estimates. Given its importance in understanding the uncertainty of the ABPEs and the impact of population components on it, it is indeed crucial to develop and provide reliable measures of uncertainty for international migration.
17. The development of migration statistics includes a rigorous process of quality assurance (QA) at all stages of data production (ONS 25/05/2023). One of the QA aspects is comparison with different sources. To aid the process of developing new methodology, the ONS could consider comparisons of their migration estimates, e.g. emigration of British nationals, with mirror statistics in other countries that are considered to have high-quality migration statistics (such as Sweden and other Nordic countries; cf. De Beer et al. 2010; Dańko et al.

[4] A long-term migrant is 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. From the perspective of the country of departure, the person will be a long-term emigrant and from that of the country of arrival, the person will be a long-term immigrant (UNdata | glossary).

2024). This could help testing the use of admin data and the assessment of the methods for producing provisional estimates.

18. The internal migration and cross-border migration are derived from the Personal Demographics Service (PDS) that is based on the registrations with a GP (ONS 23/11/2023b). The data are adjusted by using HESA data on persons moving to or leaving higher education and who are slow to update their health registration. The cross-border moves are further agreed with the National Records of Scotland and Northern Ireland Statistics and Research Agency. The adjustment by using the HESA data potentially removes an important bias in the internal migration data, especially for young working age males. However, of concern is the quality of the data as demonstrated through the linkage of the DI to the 2021 Census and CCS (ONS 01/03/2023, Figure 6). This analysis showed that the PDS has shortcomings in terms of having a matching local authority on Census-CCS records (for some age groups this matching was achieved for 75% of records). Also, the internal migration data rely on annual snapshots from the PDS with adjustments based on weekly updates. When coupled with the fact that not everyone registers with a new GP when moving, this may lead to uncertainty in internal migration especially when movements are unstable, such as during the COVID-19 pandemic restrictions.

Modelling framework

Demographic accounts framework

19. The DPM is a modelling framework for population accounts that relies on solid theoretical foundations of the Bayesian demographic accounting approach developed by Bryant & Graham (2013, 2015) and Bryant & Zhang (2018). The framework permits the estimation of coherent population accounts that satisfy the population balancing equation, including the assumption that internal migration does not change the size of the population (i.e., internal net migration is always zero). Coherency is achieved by explicitly modelling the unobservable accounts with their constraints such as population balancing equation, and it permits incorporating various dimensions (i.e., the characteristics of the population), such as age, sex, region of residence, ethnicity, and other.
20. The framework relies on Bayesian inference which provides a natural method for producing measures of uncertainty of the ABPEs and the assessment of their reliability. The use of population stocks based on administrative sources compiled every year rather than relying on updating of the census, should lead, in principle, to the removal of the uncertainty (and bias) in population estimates due to the UPC or census drift (e.g. ONS 27/06/2023a Figures 1 & 2), which is a key advantage over the current cohort-component-based approach of producing MYE.
21. There is evidence of a strong engagement of the ONS with leading academics working on the methods implemented in the DPM, including organising workshops in academic conferences. The DPM development also learnt from experiences of other statistical offices, such as Australia, New Zealand, and Italy (Blackwell et al. 2022), which helped identify key strengths and limitations of the methods. Model development would also benefit from a systematic engagement with the ABPEs users, especially those working with population characteristics for which the potential errors can be largest, such as at local authorities. An example of such engagement was a case study presented in ONS (23/11/2022).
22. The development of the DPM encountered computational challenges early in the project. This led the team to develop a novel, cutting-edge approach to estimating model parameters and producing the ABPEs. This, however, was achieved at a cost of simplifications in the demographic accounting model, which are not estimated jointly as

originally proposed in Bryant & Graham (2013, 2015) and Bryant & Zhang (2018). Briefly, the current procedure is to produce (i) approximate components of the accounts that do not have to be consistent, (ii) estimate model for demographic rates by using proxy estimates from (i); (iii) re-estimate demographic accounts individually for each local authority and use them to (iv) derive combined accounts and coherent estimates of internal migration.

23. The justification and rationale for using this approach are reasonable and especially important during the development of the modelling framework, when many models need to be tested. The rationale is also articulated in the technical documentation (Elliott & Blackwell 2023, Blackwell et al. 2022). However, in my opinion, the computational aspects should not be a major detractor from the foundations of the population accounts, such as fully coherent demographic accounts (as originally proposed in Bryant & Graham 2013), multiregional (Raymer et al. 2020) or bi-regional models (Wilson 2016). Research on foundations should go in line with the developments of the specific aspects of the model, such as new ways of smoothing data, hierarchical structure (currently not present in the DPM) and incorporation of other data sources. If indeed the computation of the joint model is not deemed possible, even with the aid of high-performance computing, the justification could be enriched by providing a comparison of the full model implemented on a small scale with the simplified approach followed by an analysis of errors, similarly to how an assessment of the computational methods and their impacts on the ABPEs has been demonstrated by ONS (18/12/2023).

The need for high-quality data

24. In general, the DPM is advertised as a flexible and adjustable platform that can accommodate a variety of population-related estimates to satisfy a variety of demands (Bryant & Zhang 2018; Elliott & Blackwell 2023). For instance, the ONS considers using the model for estimating breakdown by ethnicity, labour force status; it is also possible to provide monthly population estimates (Elliott & Blackwell 2023). However, there are at least two difficulties with this approach: the need for high-quality inputs and scalability.
25. Firstly, as has been demonstrated by the analyses carried out by the ONS, the model-based ABPEs can be sensitive to the inputs in the data (e.g., the sensitivity to the assumed precision of the MYE and SPD inputs, ONS 14/07/2022; comparison of three versions of ABPEs that utilise or not utilise 2021 Census, ONS 28/02/2023a, 28/02/2023b). As discussed later (Point 34), the inputs can be modified before being used in the model, or the model parameters can correct for data inadequacies. However, both approaches require deep understanding of the data generating processes of all DPM inputs (which the ONS generally demonstrate through their reports, e.g. ONS 2022 but also acknowledges the need for more work in this area, see e.g. Points 10 & 16) and a thorough testing within the DPM, which I understand from the communications with the ONS is work-in-progress. The ONS are well-aware of the need for a coverage benchmark and several options have been proposed and evaluated through simulation studies (ONS 2022, Elliott & Blackwell 2023, ONS 27/06/2023a). These evaluations and the transparency of the data quality assessments internally and externally to the ONS will be crucial to future deployments of the DPM and trust in the model-based ABPEs. The DPM is indeed a flexible and modifiable approach but without benchmarks to correct for biases in admin data, it may produce biased or uncertain ABPEs (acknowledged in ONS 2022). Thus, a robust quality assurance should be in place starting from data production and ending with the testing of the ABPEs.
26. The above issue is exemplified by the sensitivity analyses and comparisons of the model-based ABPEs with the official 2021 Census-based MYE (e.g., 14/07/2022, 28/02/2023a, 27/06/2023a, 27/06/2023b). ONS (28/02/2023) showed that when Census was used as input

to the model and adjustment for coverage, the differences between model-based ABPEs and 2021-Census-MYE were minimal. However, when 2021 Census was used only as a coverage benchmark (and it is the best coverage benchmark available), the ABPEs for selected LAs can differ by 1-2% from the 2021-Census-MYE and much larger differences were found for detailed breakdowns by age. These analyses demonstrated that the choice of the coverage benchmark and the model inputs can modify the model-based ABPEs, especially at the local level and for detailed characteristics.

27. These findings and sensitivity of the model-based population estimates to the assumptions have been corroborated by a study carried out on Italian national accounts (Taglioni 2019), who tested the same Bayesian demographic accounting approach of Bryant and Graham (2013). The study examined models that incorporated the hierarchical structure and were estimated jointly, unlike the models currently proposed by the ONS. This study showed that the models for population components (current Step (ii) in the DPM, Point 22) can be sensitive to the choice of the informative prior distributions for the model parameters and their choice can be crucial to the resulting population estimates. However, the accounts models (current Step (iii) in the DPM) were less sensitive to the choice of the model. Further, the study revealed that in a situation where there are differences in population size in two data sources deemed to be of high-quality (Italian 2011 census and population register), fine-tuning of the model to produce reasonable results can be difficult, especially in the context of comparing multiple hierarchical models, which, based on the documentation, has not yet been tested by the ONS. I consider it important that the ONS provides evidence of testing the DPM in extreme situations that might be encountered in the future in terms of sudden changes in data-generating processes (e.g. through changes in the legal frameworks governing administrative systems and how people interact with them) that may affect the quality of the inputs. The ONS currently provides thorough comparisons of the ABPEs based on various iterations of the DPM (e.g., 28/02/2023b, 18/12/2023b). I am also aware that the ONS has been testing various assumptions of the model, e.g. about the distributional assumptions for the data (based on unpublished documents or those still in-preparation).[5] I advise that comparisons could be presented in one document or a website, where various estimates can be compared with the 2021-Census-MYE but also between each other, ideally accompanied by the measures of errors (bias, e.g., via Mean Percentage Error, and precision, via e.g. Mean Absolute Percentage Error) across various dimensions (cf. Dańko et al. 2024).
28. The second challenge to realising the potential of the DPM and its flexibility might be the scalability due to the above-mentioned computational difficulties (also corroborated by Taglioni 2019). Bayesian computational methods are complex and may require updating, as has already been demonstrated by the ONS (18/12/2023). Introduction of new dimensions and data sources may bring additional computing cost that will be prohibitive, even in the simplified framework. Model estimation also depends on a variety of packages in open software R. While this is in principle an approach I endorse and recommend, it is also susceptible to risks such as packages not being available/maintained/compatible with other

[5] The other important tests could include goodness of fit measures via prior and posterior predictive checks, which are a typical component of a Bayesian workflow (e.g. Gabry et al. 2019). These checks include predicting the data inputs by using a model with only prior distributions, or estimated by using data inputs, potentially perturbated by random or systematic removals of portions of the data. I have been advised that such tests are being carried out by the ONS team working on the DPM development. Further, the current framework does not include hierarchical structure that permits borrowing of information across various model dimensions (as was tested by Taglioni 2019). I recommend research on the DPM is extended to test the robustness of the current modelling framework when hierarchical models are implemented for population components.

packages in the future. Therefore, in my opinion, it is important to maintain or develop within the ONS the capacity for implementing computational methods used in and around the DPM.

Migration component

29. The last point of the procedure of reconciling internal migration (Step (iv) in Point 22) is potentially risky for the quality of the final estimates because internal migration can be the largest source of uncertainty for the characteristics of the population such as age, sex and local authority (Bryant & Graham 2013; Taglioni 2019). The currently used approach utilises an iterative proportional fitting algorithm (Elliott & Blackwell 2023), which is a reliable method for reconciling demographic accounts amongst LAs. The method uses as a starting point internal origin-destination migration data derived from the PDS (Point 18), but these data do not necessarily satisfy the demographic identity. The reconciliation is carried out to the known margins (Elliott & Blackwell 2023). However, the uncertainty in the internal migration estimates is derived purely from the margins and not the estimation procedure. Of concern might also be the bias in the data inputs as exemplified by the mismatch of the PDS (via DI) and 2021-Census-CSS (ONS 01/03/2023, see also Point 18).
30. Moreover, in the current approach, migration to and from each of the LAs is modelled jointly, as a single in- or out-flow to or from the local authority. Since the estimates of international migration are derived from a variety of sources (Point 14) and the DPM requires a careful setting of informative priors, it might thus be challenging to formulate a prior for the migration component as the uncertainty of the estimates may depend on the prior. I suggest that a thorough testing of the sensitivity of the ABPEs to the choice of this prior is performed and documented.
31. Further, a lack of a high-quality benchmark for the internal migration estimates to adjust for coverage issues and delays in reporting, together with any potential biases in the international migration data may lead to errors in the ABPEs that could potentially be similar to the intercensal drift, for example, if the benchmark based on the 2021 Census is continued to be used in the future. This may affect selected local authorities or sub-populations (e.g. mobile working-age persons, those working from home), see Point 35.

Documentation and reproducibility

32. As part of the assessment, I analysed the documentation of a part of the computer code used in producing the DPM population estimates that was made available by the ONS. The documentation is transparent with examples of how to use the R packages developed by the ONS team. I am aware of a detailed log file where issues with the code and the model are being reported. While certain aspects of the code seem to be indeed in development, the provided documentation assures that the model can be tested internally and potentially also by external stakeholders. As mentioned before, my opinion is that the ONS should develop and maintain the capacity to implement the future changes in the DPM methodology, computational methods and their implementation. This will ensure the new admin-based population estimation system is sustainable in the future and the ONS is on track to achieving this with the computer code they are developing.
33. A development by the ONS team of the R packages that implement the DPM and let users produce their own ABPEs are crucial for ensuring the reproducibility of the results and, thus, further testing of the DPM and ABPEs robustness. The current versions of the packages permit the estimation for selected local authorities and production of publishable documents with the analyses of the results that are generated by the package. A potential issue with reproducibility, especially if the packages are to be shared with partners external

to the ONS, is that the exact data inputs may not be directly accessible due to data sharing agreements in place. These issues can be mitigated by an interactive R package that I understand is being developed by the ONS. This package could contain a toy model that would demonstrate the workflow of the DPM and permit testing of (some of) the model assumptions, perhaps by using synthetic (not real) data. This would also benefit the communications of the estimates especially to stakeholders interested in a more detailed understanding of the model (such as academics, local authorities, government departments).

Uncertainty

34. The uncertainty of the data inputs is directly used in the DPM through the model parameters (e.g., dispersion parameters of population component rates, Eq. 4.4 in Elliott & Blackwell 2023, are informed by the ONS estimates of standard errors). They are therefore crucial for providing the uncertainty measures of the ABPEs derived from the DPM. There are two avenues of quantifying uncertainty within the DPM: (i) via assumed known parameters (variance, dispersion) where this uncertainty is derived for each source externally, such as for the 2011 Census base (Point 8); or (ii) via estimable parameters that capture variability of the data. The latter approach requires formulating prior distributions ("priors") for the uncertainty parameters. Such priors may be uninformative (i.e., driven purely by data) or informative, e.g. based on the externally derived uncertainty and/or informed by demographic expertise. The Bayesian demographic accounts framework requires informative priors (Bryant & Zhang 2018, Taglioni 2019). A potential risk with the currently used option (i) is that the uncertainty constructed for the data inputs externally is propagated in the DPM and so the quality of the final estimates depends on that uncertainty assessment. A need for a sustainable framework for quality measures of the DPM inputs and administrative data has been acknowledged by the ONS (ONS 2023a, ONS 2023b: table A1). I fully support such a pledge. I further recommend that the framework is extended to the ABPEs derived from the DPM. Further, as part of standard model checks (Gelman et al. 2013; Bryant & Zhang 2019) sensitivity analyses should be carried out testing the sensitivity of the ABPEs to the assumed uncertainty parameters for each of the data inputs.
35. The ONS has published an example of such sensitivity analysis for a synthetic local authority (ONS 14/07/2022), where they demonstrated that the model-based ABPEs can be closer to SPD or to MYE, depending on the relative value of the uncertainty (precision) parameter. It is important to include such sensitivity tests in the workflow for the final estimates for all local authorities as the quality of inputs may vary between and within sources over time, age, or across local authorities. Case in point, it is acknowledged that some LAs have "time lags in the accuracy of administrative data" due to high levels of migration, high percentage of rental houses or being urban areas (ONS 28/02/2023d). Further, the sensitivity analysis should specify the meaning of the assumptions on precision, e.g., in terms of coefficients of variation around the means, rather than stating that one data source is more precise than the other.
36. Because of these differences between local authorities, the future versions of the DPM may also permit differentiating the precision of data sources between the LAs, depending on their characteristics. This might be done by creating a typology of LAs that share common characteristics and an introduction of hierarchical components in the DPM that capture these characteristics. In this context, it is relevant to engage with the stakeholders at local levels (local authorities) to elicit any insights that they may offer in terms of characteristics of the population that may not be captured well by the administrative sources. The ONS has

carried out such consultation with 14 Local Authorities, where the results from the DPM were produced and compared with the 2021 Census (ONS 23/11/2022). This exercise prompted a revision in the sub-model of the DPM for migration, as specific age groups were not estimated as expected. If the DPM-based ABPEs were to become official population statistics to be used for policymaking at a local level, it might be of value to consult the estimates and any major changes and updates in the methodology with those key stakeholders, explaining how these updates may affect the estimates and what the risks and benefits of the model updates are. Feedback from stakeholders may also lead to future revisions of the methodology.

37. As it is presented currently, the DPM does not account for all sources of uncertainty. For example, the uncertainty around ABPEs produced in 2020 (ONS 27/07/2020) was based on the variability in ABPEs for "similar" local authorities scaled to the 2011 Census – but without acknowledging the uncertainty of the census estimates (Point 8). Another example is uncertainty related to probabilistic linkage used in creating the DI and then SPD – a key input to the DPM (Point 10) does not seem to be reflected in the uncertainty measures in the DPM. Simultaneously, criteria for the quality of the estimates in terms of the width of the confidence/credible interval relative to the estimated population size are being implemented (ONS 2023c). This creates a potential risk of underestimating the uncertainty of the ABPEs and, thus, giving users a false sense of precision of the estimates. I recommend that reported model-based ABPEs and their precision are accompanied by a statement on the potential sources of uncertainty that are unaccounted for and, where possible, an assessment of their importance in a given situation.
38. The measures of uncertainty, of both the DPM inputs as well as its outputs, can pose a challenge for interpretation by the various stakeholders. The ONS has demonstrated a good understanding of the need for communicating uncertainty that follows the recommendations of the Office for Statistics Regulation (2022). One of the approaches to a better representation of uncertainty (and more widely, data quality) is an assessment in terms of the risks of potential cost of under- or over-estimating population counts/rates in a given source or by the DPM. This would also help identify which population characteristics may be of greater concern to stakeholders (cf. Bijak et al. 2019) and could inform where more research is needed to better understand the source of the uncertainty in data inputs.

Concluding remarks

39. Overall, the DPM lays solid foundations for producing population estimates for England and Wales based on administrative data. The key advantage of the DPM over the current cohort-component-based approach is that the drift related to updating data from the census and increasing uncertainty can be reduced and more timely estimates can be produced. This, however, depends almost entirely on the data inputs to the model and the assessment and understanding of their quality. This assessment is crucial because the relative differences in quality between data sources can influence the model-based ABPEs and their quality in terms of uncertainty or potential biases.
40. The key aspect of reliable ABPEs based on the DPM and administrative data is understanding and measuring the under- and overcoverage of the various data sources and an implementation of a long-term strategy of providing a high-quality benchmark for the administrative data. One of such benchmarks is census, but it is available only for 2011 and 2021. Various other options, such as coverage surveys, addresses register and population register, have been proposed and studied by the ONS. In the case of a lack of a reliable benchmark that can be used to adjust data inputs or correct for data inadequacies in the

model, there is a risk that the admin-based ABPEs will still suffer from an error, which may have a similar nature to the intercensal drift, for example, in a situation when coverage adjustments are based on the 2021 Census and are extrapolated for the future releases of the ABPEs.

41. In my opinion, more research is also needed to better understand the nature of errors (biases and uncertainty) in the data sources that are used as inputs to the DPM. This should be done in consultation with a variety of stakeholders and through their engagement with the model development (such as through a demonstrative R package). The key stakeholder are local authorities, amongst which there are large differences in how admin data capture populations. Such investment may help in the DPM development and ensure that biases and uncertainty in the ABPEs can be reduced.
42. The differentiation between bias and accuracy that can be present in the data can be built into the established ONS procedures in data quality assessment, such as those based on the European Statistical Systems and QQI. This would be in line with the recommendations developed regarding the theoretical quality standards for the future population estimates in terms of bias and variance (ONS 2023c). An approach similar to the one in ONS (2023c) of considering bias and accuracy separately and in each of the data inputs to the DPM, could be developed. In fact, this differentiation is being made, for example, by studying overcoverage and undercoverage (i.e. bias) of the SPD (Law et al. 2022; 2023) and, in another study, developing methods for generating measures of uncertainty for the SPD (ONS 27/07/2020). However, such analyses seem to be unrelated to each other, whereas their outcomes potentially constitute key inputs to the DPM. An assessment of errors could include analysing the data generation process in terms of what may cause a systematic error (e.g. through under- or overcoverage of sub-populations, persons systematically not interacting/delaying interactions with an admin system, admin systems being focused on documents rather than persons, etc), and what can lead to a non-systematic error (e.g. where mis-classifying a person as a resident or non-resident is equally likely, mis-classification in linkage). An example of such a framework is a Total Survey Error framework (e.g. Groves and Lyberg 2010), which, while not directly applicable, provides an overview of how errors in admin data could be described. I appreciate that situations where a distinction if a given exclusion rule or data collection mechanism generates bias or accuracy is not possible and, thus, I recommend that further research is carried out to better understand those mechanisms following good practices already established at the ONS through, for example, simulation studies and clerical checks of samples of data (cf. Law et al. 2023).
43. The model is generally well-described from a technical point of view (bearing in mind that it is still in development) and there is evidence that it has gone through testing or is being tested as is a norm in the Bayesian inference workflow. Various aspects, such as those pointed out in this report, or issues related to the current ONS approach to publishing documentation, such as results from subsequent updated models not being directly compared or comparable, minor inconsistencies in reporting, varying levels of technicalities, could be better presented in a more comprehensive way, with a clear structure related to the data inputs, modelling framework, technical assumptions, computational methods, model testing and analysis of the outputs, all in one document, dashboard or a website. For instance, subsequent versions of the DPM could be released in a similar fashion as the SPDs, which are assigned a version. All aspects of the above are already available but may require updating or are in preparation. A resulting coherent documentation will be crucial for the ABPE/DPM project to be sustainable in the future.

References

Bijak, J., Disney, G., Findlay, A. M., Forster, J. J., Smith, P. W., & Wiśniowski, A. (2019). Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting*, *38*(5), 470-487. https://doi.org/10.1002/for.2576.

Blackwell L., Elliot D., and Bryant J. (2022), UK Statistics Authority, Methodological Assurance Review Panel working paper EAP174, Integrated statistical design for the transformed population and social statistics system - Bayesian methods for demographic estimation.

Bryant, J. R., & Graham, P. J. (2013). Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources. Bayesian Analysis, 8(3), 591-622. DOI: 10.1214/13-BA820

Bryant, J. R., & Graham, P. (2015). A Bayesian approach to population estimation with administrative data. Journal of Official Statistics, 31(3), 475-487. https://doi.org/10.1515/jos-2015-0028.

Bryant, J., & Zhang, J. L. (2018). Bayesian demographic estimation and forecasting. Chapman and Hall/CRC. https://doi.org/10.1201/9780429452987.

Champion T. (2024) Analysing Unattributable Population Change (UPC): An unpublished research note by Tony Champion, Newcastle University, 09/04/2024.

Dańko MJ, Wiśniowski A, Jasilionis D, Jdanov DA, Zagheni E, (2024) Assessing the quality of data on international migration flows in Europe: The case of undercounting, Migration Studies, Volume 12, Issue 2, June 2024, mnae014, https://doi.org/10.1093/migration/mnae014

De Beer, J., Raymer, J., Van der Erf, R., & Van Wissen, L. (2010). Overcoming the problems of inconsistent international migration data: A new method applied to flows in Europe. *European journal of population= Revue europeenne de demographie*, *26*(4), 459.

Elliott D., and Blackwell L. (2023), 10 February 2023, UK Statistics Authority, Methodological Assurance Review Panel working paper EAP193, Progress on an integrated statistical design for the transformed population and social statistics system: Bayesian methods for demographic estimation.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. Journal of the Royal Statistical Society Series A: Statistics in Society, 182(2), 389-402.

Gelman A. (2008) Objections to Bayesian statistics. Bayesian Analysis 3 (3) 445 - 449, September 2008. https://doi.org/10.1214/08-BA318.

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. Public opinion quarterly, 74(5), 849-879. https://doi.org/10.1093/poq/nfq065.

Law E., Hammond C., Linton M., Ismail Z., Davies S., Shipsey I. (2022), September 2022 UK Statistics Authority, Methodological Assurance Review Panel working paper EAP184, SPD Estimation Options.

Law E., Large A., Hammond C., Linton M. (2023), September 2023 UK Statistics Authority, Methodological Assurance Review Panel working paper EAP202, Population stock estimates using linked administrative data and a coverage survey – a case study for 2021 and future directions.

Office for National Statistics (ONS), published 16 July 2012, National Archive, Explaining the Difference between the 2011 Census Estimates and the Rolled-Forward Population Estimates [ARCHIVED CONTENT] UK Government Web Archive - The National Archives.

Office for National Statistics (ONS), released December 2012, Confidence Intervals for the 2011 Census [ARCHIVED CONTENT] Confidence intervals - ONS (nationalarchives.gov.uk)

Office for National Statistics (ONS), released 30 November 2017, ONS methodology working paper series no. 10 – Methodology for measuring uncertainty in ONS local authority mid-year population estimates: 2012 to 2015

Office for National Statistics (ONS), released 27 July 2020, ONS website, Admin-based population estimates and statistical uncertainty - Office for National Statistics (ons.gov.uk)

Office for National Statistics (ONS), released 16 April 2021, ONS methodology working paper series no. 23. Using statistical modelling to estimate UK international migration - Office for National Statistics (ons.gov.uk)

Office for National Statistics (ONS), released 14 July 2022, ONS website, Dynamic population model for England and Wales - Office for National Statistics

Office for National Statistics (ONS), released 5 September 2022, ONS website, article, Rebasing and reconciliation of mid-year population estimates following Census 2021, England and Wales: 2022

Office for National Statistics (ONS) (2022), October 2022, UK Statistics Authority, Methodological Assurance Review Panel working paper EAP182, Evaluating Statistical Quality in the Demographic Index Demographic Index v2.0 and Methodology Division Quality Assessment.

Office for National Statistics (ONS), released 23 November 2022, ONS website, article, Dynamic population model for local authority case studies in England and Wales: 2011 to 2022

Office for National Statistics (ONS) (2023a), UK Statistics Authority, Methodological Assurance Review Panel working paper EAP190, Quality issues related to linkage of administrative data.

Office for National Statistics (ONS) (2023b), UK Statistics Authority, Methodological Assurance Review Panel working paper EAP191, Quality focus for population and migration statistics transformation.

Office for National Statistics (ONS) (2023c), UK Statistics Authority, Methodological Assurance Review Panel working paper EAP189, Bias and Variance quality standards for 2023 recommendation.

Office for National Statistics (ONS), released 1 March 2023, ONS website, article, Understanding quality of linked administrative data sources in England and Wales, using the 2021 Census - Demographic Index linkage

Office for National Statistics (ONS), released 28 February 2023a, ONS website, methodology, Dynamic population model, improvements to data sources and methodology for local authorities, England and Wales: 2011 to 2022

Office for National Statistics (ONS), released 28 February 2023b, ONS website, article, Admin-based population estimates: provisional estimates for local authorities in England and Wales, 2011 to 2022

Office for National Statistics (ONS), released 28 February 2023c, ONS website, article, Understanding quality of the Statistical Population Dataset in England and Wales using the 2021 Census – Demographic Index linkage.

Office for National Statistics (ONS), released 28 February 2023d, ONS website, article, Developing Statistical Population Datasets, England and Wales: 2021

Office for National Statistics (ONS), released 3 March 2023, ONS website, article, Administrative sources used to develop the Statistical Population Dataset for England and Wales: 2016 to 2021.

Office for National Statistics (ONS), released 25 May 2023, ONS website, Admin-based long-term international migration estimates QMI

Office for National Statistics (ONS), released 25 May 2023, ONS website, Admin-based long-term international migration estimates QMI

Office for National Statistics (ONS), published 01 June 2023, ONS website, methodology, Measuring uncertainty in international migration estimates

Office for National Statistics (ONS), released 27 June 2023a, ONS website, article, Admin-based population estimates: updated estimates for local authorities in England and Wales, 2021 to 2022

Office for National Statistics (ONS), released 27 June 2023b, ONS website, methodology, Dynamic population model, improvements to data sources and methodology for local authorities, England and Wales: 2021 to 2022

Office for National Statistics (ONS), released 23 November 2023a, ONS website, statistical bulletin, Population estimates for England and Wales: mid-2022

Office for National Statistics (ONS), released 23 November 2023b, ONS website, methodology, Population estimates for England and Wales, mid-2022: methods guide

Office for National Statistics (ONS), released 23 November 2023c, ONS website, methodology, Quality and methodology information (QMI) for Census 2021

Office for National Statistics (ONS), released 18 December 2023a, ONS website, methodology, Dynamic population model, improvements to data sources and methodology: local authorities in England and Wales, mid-2021 to mid-2023

Office for National Statistics (ONS), released 18 December 2023b, ONS website, article, Admin-based population estimates: local authorities in England and Wales, mid-2021 to mid-2023

Office for National Statistics (ONS), released 29 December 2023, ONS website, methodology, Mortality statistics in England and Wales QMI - Office for National Statistics

Office for National Statistics (ONS), released 23 February 2024, ONS website, methodology, Births Quality and Methodology Information

Office for National Statistics (ONS), released 3 May 2024, ONS website, methodology, Provisional long-term international migration estimates: technical user guide

Office for Statistical Regulation (OSR), September 2022, Approaches to presenting uncertainty in the statistical system (statisticsauthority.gov.uk)

Raymer, J., Bai, X., Liu, N., & Wilson, T. (2020). Reconciliation of Australian demographic data to study immigrant population change across space and time. *Spatial Demography*, *8*, 123-153.

Robert, C., & Ntzoufras, I. (2012). Bayesian modeling using WinBUGS. CHANCE, 25(2), pp. 60–61. https://doi.org/10.1080/09332480.2012.685377.

Taglioni, Ch. (2019) Bayesian hierarchical modelling for population size estimation: application to Italian data. PhD thesis, 1 October 2019, University of Padua.

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., ... & Yau, C. (2021). Bayesian statistics and modelling. Nature Reviews Methods Primers, 1(1), 1.

Wilson, T. (2016). Evaluation of alternative cohort-component models for local area population forecasts. *Population Research and Policy Review*, *35*, 241-261.

List of acronyms

ABMEs	Admin-based migration estimates
ABPEs	Admin-based population estimates
CCS	Census Coverage Survey
DI	Demographic Index
DPM	Dynamic Population Model
GP	General Practitioner
HESA	Higher Education Statistics Agency
IPS	International Passenger Survey
LA	Local Authority
LTIM	long-term international migration
MYE	(Population) Mid-year estimates
ONS	Office for National Statistics
PBE	Population Balancing Equation
PDS	Personal Demographics Service
QQI	Quantitative Quality Indicators
RAPID	Registration of Population Interactions Database
SPD	Statistical Population Dataset
UPC	Unattributable population change