Population and migration statistics transformation in England and Wales, a quality strategy

This report provides information about the quality strategy underpinning the future population and migration statistics system in England and Wales.

Table of contents

1 . Main points

- The Office for National Statistics (ONS) is transforming the population and migration statistics system for England and Wales, making use of the best available data sources, using robust and innovative methods.

- The system will deliver timely, coherent and accurate statistics, providing new insights to support a better understanding of the population; critical for effective decision making to improve people's lives.

- The use of new sources (including administrative data) and methods must be supported by a robust quality strategy, to ensure that outputs meet the needs of users and users understand the strengths and limitations.

- This report provides details about the strategy, including the principles on which it is based, and the framework for assessing and managing quality; it includes the ONS's current work and key areas for future development.

- The strategy will evolve, as ONS advances from research and experimental statistics, to producing National Statistics using the future system; influenced by user feedback.

2 . Introduction

The Office for National Statistics (ONS) is transforming the population and migration statistics system, using a range of data sources from across government and the public sector. This is to deliver more timely and frequent statistics about the population down to local levels, as shown in our [Overview of population and migration statistics transformation](). The future system must be supported by a robust quality strategy, which covers the quality of the data sources (input quality), data processing (process quality) and the resulting quality of the final statistical outputs; ensuring they meet user needs (output quality). This report provides information about the strategy and our future plans.

3 . The quality principles

Underpinning the quality strategy for the future population and migration statistics system, are the following high-level principles (covering the input, process and output stages):

- statistical outputs are relevant against priorityuser needs

- the most appropriate data sources are used to deliver the statistical outputs

- strengths and limitations ofdata sources are understoodagainst their use, documented and clearly communicated to users

- strengths and limitations ofthe data sources are accounted for in design decisions,concerning the use of the sources

- changes in quality, through theintegration and processingof data, are quantified and reported on

- statistical methodsused to deliver the statistical outputs are robust, based on best practice and have been endorsed and quality assured by experts in the field

- methods for quality assessment, improvement and reporting, draw on best practice and input from experts in the field

- statistical outputsare assessed against anunderstanding of total errorfrom the data sources and the processing of the sources

- statistical outputs are assessed through comparisons with other sources, comparisons over time, internal coherence, against known demographic relationships and by drawing on expert scrutiny

- quality isassessed on a continual basis, to ensure changes in quality are understood, accounted for and reported on via feedback loops, including with data suppliers

- flexibility exists in the design of our statistics, by avoiding over-reliance on a single data source and ensuring that inputs have complementary sources of known quality, if they are required

4 . Quality framework

We define quality to mean that the statistical outputs fit their intended uses, are based on appropriate data and methods, and are not materially misleading ([Code of Practice for Statistics](...)). The quality strategy for the future population and migration statistics system focuses on statistical quality throughout the statistical journey, from data collection to use of our statistics in decision making. Understanding and communicating quality is essential to assure and guide users of the Office for National Statistics (ONS) statistical products.

The strategy assesses quality against three stages:

- input quality (the quality of data sources used in the future system)

- process quality (quality through the data processing stage)

- output quality (the quality of the resulting statistics)

The quality assessment at the input quality stage not only informs whether a source is suitable and its use, but also the development of statistical methods and the necessary processing of the source (for example, whether and how it will be integrated with other sources and whether editing or imputation is needed). An understanding of quality at the input and process stages then relate directly to the quality of the statistical outputs. There is a feedback loop across the stages, for example, quality issues identified in the outputs can inform future improvements to the data sources, statistical methods and data processing.

Across the quality stages, the ONS will ensure the right tools and capabilities are in place and that quality is reviewed on a continual basis, as part of our [ONS Statistical Quality Improvement Strategy](...).

Input Quality

The transformed system will use a range of data sources, including administrative data (information collected by government and/or other organisations primarily for administrative purposes, such as, registration, transaction and record keeping), other commercial data (for example, mobile phone data) and survey data. For ONS surveys, the ONS has control over the content and design; however, administrative data have not been collected for ONS statistical purposes and therefore require other methods of assessment.

Administrative and commercial data quality

The following dimensions are used to assess the quality of administrative data sources and commercial data:

Relevance and integrability

The extent to which the source meets the intended needs, including how well it covers the population of interest and the required attributes (for example, age and ethnicity). The extent to which the source can effectively be integrated into the statistics system; including validity (the extent to which the data conform to expected format, type, range) and the presence of high-quality linkage variables, if required.

Time-related dimensions

Coherence over time (changes in concepts, definitions and coverage) and timeliness (that is the period between the date to which the information relates to, and the date the data are available to the ONS).

Accuracy and reliability

The degree to which data describe what they were intended to measure. The completeness (levels of missingness and population coverage), uniqueness (degree to which there is no duplication of records). Plausibility and consistency (values do not conflict with other values within a dataset or across datasets).

Delivery and clarity

The confidence data will arrive as and when needed, including the robustness of data sharing agreements, and the relationships with data suppliers. The completeness and clarity of metadata that describe the data source. The availability of related data sources to support, complement or substitute; reducing the reliance on any one source.

Information about the administrative data sources used in the research to transform the population and migration statistics system, along with information about their quality is available in our Data source overviews.

An important part of the assessment process is the relationship and communication with data suppliers. This relationship is important to build an in depth understanding of the data sources (clarity) and to ensure the right agreements are in place to secure data as and when needed. Good communication also means that the ONS is consulted on changes to the sources and can feedback on areas for quality improvement.

The ONS uses a variety of communication approaches as part of the quality strategy, including:

- quality working groups with data suppliers

- secondments into government departments

- interviews and shadowing those involved in data collection and processing

- meetings with key areas within supplier departments (data managers, IT specialists, operational teams and data analysts)

The ONS will also draw on the Admin Data Quality Question Bank, which is scheduled to be published in summer 2023. The question bank includes a tested question set for obtaining quality information about the data from the supplier and also provides an administrative data quality checklist for analysts.

In terms of the methods and approach to assessment, the strategy draws on best practice from the [Office for Statistics Regulation (OSR)](), the Government Statistics and Social Research Services, and international approaches and developments. This includes the use of the [Quality of Admin Data in Statistics Framework](), the [Quality Assurance of Administrative Data toolkit]() and our [cataloguing error in administrative and alternative data sources]() publication.

We will also continue our research [Exploring the quality of administrative data using qualitative methods](), which has provided useful insight into how different groups of the population interact with administrative data sources.

Survey data quality

The ONS will continue to draw on the wealth of experience it has in ensuring the collection of high-quality survey data, supported by quality assessment and improvement activities. This includes:

- sampling design

- questionnaire development

- fieldwork practices and training

- validation and quality assurance of data through the data collection stages

- ensuring the design of the surveys is relevant against the user needs for the transformed population and social statistics system

Process quality

Once data are received, they are then integrated with other sources into the statistical system and quality issues are addressed (as identified at the "Input Stage"). Such issues may include concept misalignment between the administrative data sources and the required statistical concepts and definitions; coverage error and measurement errors.

Important processes include the engineering of data to the required structure and harmonised standards, data validation, linkage, editing and imputation, data modelling and estimation. The aim of the processing is to ensure the production of accurate and timely statistics, through a good understanding of the properties, strengths and weaknesses of the source data.

Sound methods

The processing stage must be underpinned by the best available methods and recognised standards for producing statistics from administrative and survey data sources. The methods that have been and continue to be developed will draw on international best practice and will be reviewed by an external Methodological Assurance Review Panel (MARP). Further information about the methods and MARP can be found on our[Methodology and quality strategy]() page.

Processing and assurance

As data are processed using sound methods, checks are in place to assess changes in data quality as the data go through the production stages. This includes data visualisation techniques. We are also building [Reproducible Analytical Pipelines (RAPs)](#) for running our processes and applying our methods to ensure our outputs are reproducible, adaptable and sustainable, using automation and good software engineering practices.

The RAPs also allow consistent, auditable and high-quality assessment and interrogation of specific quality metrics. The metrics support a process of ongoing review and development to ensure that the processes are working effectively.

Our approach aligns with the [Office for Statistical Regulation best practice](#) and [Government Analysis Function guidance on RAPs](#). It is supported by the [ONS Digital and Technology Strategy (PDF, 982KB)](#), which describes how continuous improvement and automation will be at the heart of the ONS service provision. The result is increased efficiency and transparency of processes, which will enhance trust in the resulting analysis for producers and users.

Data integration and linkage

The transformed system integrates multiple data sources to ensure the resulting statistics are inclusive of the whole population and use the best combination of data available. An important process for integrating data is data linkage. This is particularly challenging for administrative data (when compared against surveys), as a result of the availability of variables to check the match should really be a match. For more information see the [UK Statistics Authority, Quality issues related to linkage of administrative data paper, EAP190 (PDF, 200KB)](#). Development of our methods for linking administrative data and assessing the quality of linkage is an important part of the quality strategy.

The ONS assesses the quality of linkage through an understanding of the quality of the variables used to link data together, and an assessment of the linkage process and outputs. This includes estimation of false positives (records that have been linked in error) and false negatives (records that were not linked when they should have been). It also includes comparisons of the distribution of the characteristics of linked and unlinked records (for example, by age, sex and other characteristics) to understand potential biases.

The framework used by the ONS for linking data is called the Reference Data Management Framework (RDMF). The RDMF enables the ONS to separate the data linkage function (where identifiers such as name and address are used to link datasets) from subsequent data processing (where de-identified linked data is then used).

The RDMF is used to construct the Demographic Index (DI). The DI integrates education, health, and tax and benefit administrative data to provide a composite data source of the population interacting with administrative data sources. The DI is a building block for the [Statistical Population Dataset (SPD)](#) that is fed into the [Dynamic Population Model (DPM)](#), to produce admin-based population estimates. The SPD approximates the usual resident population of England and Wales, using integrated administrative data sources.

As part of the quality strategy, the ONS has proposed a range of metrics to understand the statistical quality of the DI, including linkage quality ([UK Statistical Authority, Evaluating Statistical Quality in the Demographic Index paper, EAP182 (PDF, 549KB)](#). The quality of the DI and the SPD is also being assessed through linkage to the Census 2021 and Census Coverage Survey ([A linkage project between the 2021 Census and Census Coverage Survey to the Demographic Index: Rationale and Research Questions, EAP192 (PDF, 523KB)](#). This provides important quality information about differences between the population captured in the 2021 Census and the population in the DI and SPD, as found in our [Understanding quality of linked administrative data sources in England and Wales, using the 2021 Census - Demographic Index linkage article](#). This information will support improvements to data linkage and the methods used to produce the SPD.

Administrative data are also linked longitudinally as part of our research, including for statistics on internal and international migration. We have an [error framework for longitudinal administrative sources](#), which will be used to understand the quality of longitudinally linked administrative data.

It is important that our linkage methods are inclusive, without introducing biases for groups of the population for which accurate linkage is more challenging. We will therefore continue our work to develop linkage methods for more complex population groups, building on the work of our [Refugee Integration Outcomes (RIO) data linkage pilot](#) that uses innovate approaches to link data for refugees.

Processes for dealing with missingness and making decisions between sources

It is important to account for missingness in the data sources to improve accuracy. This may be achieved through the application of imputation methods in the processing stage. The methods will be assessed for quality to ensure they are improving accuracy, preserving the distributions of the true data values and delivering results that are plausible and consistent. The ONS is working to overcome the challenges of missingness and conflicting records, which includes our research from the Methods for producing multivariate population statistics using administrative and survey sources paper, EAP186 (PDF, 353).

We are also exploring new methods for comparing and making decisions between data sources, including estimating error, where there are multiple data sources each including a variable which measures the same (or a nearly identical) concept. This includes techniques such as structural equation modelling (SEM) and multiple imputation latent class (MILC) modelling, which can take account of error and factor uncertainty into final estimates.

Output Quality

It is important that the statistical outputs from the future population and migration statistics meet user needs, are well understood, and that users understand the strengths and limitations of the statistics. The European Statistics System (ESS) output quality dimensions are used as part of the strategy to assess output quality. These are described below.

Relevance

The degree to which the statistical outputs meet current and emerging user needs.

Accuracy and reliability

The closeness between an estimated result and the unknown true value, and how reliable these are over time and geography.

Timeliness and punctuality

The lapse of time between publication and the period to which the data refer, and the time lag between actual and planned publication dates.

Accessibility and clarity

The ease in which the data user can access and understand the data they are interested in.

Coherence and comparability

The degree to which data can be compared over time, region and domain. The degree to which data that are derived from different sources or methods, but refer to the same phenomenon, are similar.

Relevance

To ensure that outputs from the transformed system are relevant, the ONS will continue to work across the user community. This includes through various events and forums, and by invited feedback on our programme of research and statistical outputs.

In June 2023, the ONS will launch a consultation on the future of population and migration statistics in England and Wales. Responses to the consultation will provide us with evidence about how our proposals for the future population and migration statistics system meet the needs of users, and the confidence users have in these proposals based on our research to date. This information will be used to ensure our research and outputs remain relevant.

Improving accuracy, timeliness and coherence

The ONS research into the use of administrative and other data sources to produce population and migration statistics is demonstrating our ability to deliver more frequent, timely, inclusive and responsive statistics, which better meet the needs of users. For example, the development of a Dynamic Population Model (DPM) has provided a coherent statistical framework for more timely population statistics to meet core user needs.

The population statistics produced from the DPM, using multiple data sources, will also sustain a better level of accuracy over the decade than is possible with the existing system. Information about the research, including statistics on the population and migration, housing and households, and our work on longitudinal analysis and outcomes is available on our Research outputs using administrative data page.

Assessing accuracy and quality standards

To assess the accuracy of the transformed statistics, we are working to understand two things: firstly, what is the statistical quality of the estimates from our current system (based on the census and Mid-year population estimates QMI), that users are accustomed to; secondly, how do we develop methods to quantify the statistical quality of the outputs from the transformed system to inform users of the certainty of these outputs and allow a comparison.

To support the first, a set of quality standards for population size estimates has been developed. These quality standards refer to the level of statistical quality that we are aiming to achieve from the transformed system. We consider quality in terms of variance (how precise an estimate is) and bias (whether we are systematically under or over-estimating our values). We have based these standards on those achieved by Census 2021 and the mid-year estimates (MYE) based system. We looked at the statistical quality of outputs across the decade between 2011 Census and the year before Census 2021, to understand both the high statistical quality of census outputs and the decline in quality as we move further away from census base year.

Our standard therefore is based on the quality estimated for the MYE 2016, as that reflects the average quality standard over the decade, mid-way between the two censuses. Further information on the quality standards can be found in the UK Statistics Authority Bias and variance quality standards for 2023 recommendation, EAP189 (PDF, 223KB).

For the second, we are also developing methods to quantify the statistical uncertainty of the new statistical outputs. This will allow users to consider how they prioritise frequency and geographic granularity against accuracy and will inform how the ONS develops the outputs in the future. Some of this work has already begun: the ONS has outlined its proposals for measuring uncertainty in admin-data based international migration estimates. This is a new and innovative area of methods development.

We have also recently released our initial methods for calculating uncertainty for DPM estimates at the local authority level. This is an approximate method which, as it is novel, needs refinement, but indicates that the DPM based population estimates have a more consistent level of quality over the ten-year period than current mid-year estimates. Census and DPM-based estimates cannot be directly compared with each other, owing to fundamental differences in how estimates are produced. Each has their own advantages and drawbacks, which are explored in greater detail in our Dynamic population model, improvements to data sources and methodology for local authorities, England and Wales: 2021 to 2022 paper, due to be published 27 June 2023.

Our future work will include expanding these quality standards for population estimates to cover other statistics about the characteristic of the population.

Quality assurance of the outputs (supporting coherence and comparability)

In addition to the checks carried out in the processing stage, we use a number of approaches to quality assure the population statistics. This includes comparisons with other alternative sources, including the Census 2021 and survey data. It also includes the monitoring of trends in the components and for key demographic indicators, such as, mortality, childbearing, migration and sex ratios. We have developed a dashboard that allows real-time assessment and visualisation of trends, which is monitored by data experts and demographers.

The dashboard allows us to visualise and compare recent and historic patterns, so that we can identify unexpected changes in the data. We are also exploring new and innovative "signal" data to include in the dashboard that provides further intelligence about population change at local level, including mobile phone usage and energy consumption, such as, electricity or gas data. This information not only supports quality assurance of the results, but the sources may also prove valuable for inclusion into our models for population estimates in the future.

We also draw on expert review of the methods and estimates, and local area intelligence, including our work with local authorities. This has been invaluable in developing the DPM and understanding what other data sources might support the quality assurance or development of our outputs. For this purpose we have launched a local population statistics insight feedback framework, as described in our Receiving user insights on local population levels and change, England and Wales: August 2022 article. The framework enables users of population statistics to provide feedback at local authority level and suggest sources for us to better understand the quality of our estimates.

Accessibility and clarity

The ONS will work with users to ensure our statistics can be easily accessed; users can retrieve the information they need quickly and in the format they require. To achieve this, we will build on the accessibility and successful outputs tools used for Census 2021 data, which includes creating a custom dataset, as shown in our New ways to access Census 2021 data news and insight page.

We will continue to publish information on the data sources, methods and processes, as our work progresses, to ensure users understand how our statistics have been produced and their quality. We will listen to and take on user feedback to continue to improve our approaches and the way we communicate information.

5 . Future plans

The Office for National Statistics (ONS) will continue to work on developing a quality-driven, statistical framework for combining administrative and survey data to produce population statistics. This includes the work to further assess the quality of individual and linked administrative datasets that are used as part of the Dynamic Population Model (DPM) framework for producing admin-based population statistics. Our work will draw on international best practice and developments.

We will continue to develop methods to quantify the statistical uncertainty of the new outputs, including our work to set out quality standards for population and characteristics statistics.

We will continue to publish quality information about the data sources, our processing and the resulting statistical outputs, so that users understand the quality of the transformed population and migration statistics. This includes a case study on the quality of the Statistical Population Dataset (SPD), scheduled to be published in summer 2023.

We will continue to engage with users, including through the feedback as part of the consultation on the future of population and migration statistics in England and Wales. User feedback will support the development of the quality strategy.

We will further develop the structures around quality management, guided by the ONS Statistical Quality Improvement Strategy, as we move from research towards producing accredited National Statistics under the future population and migration system.

6 . Cite this methodology

Office for National Statistics (ONS), released 26 June 2023, ONS website, Methodology, [Population and migration statistics transformation in England and Wales, a quality strategy](#)