Population and Social Statistics Transformation – an update

Sarah Henry – Director of Methods and Quality
Louisa Blackwell – Head of Integrated Statistical Design
Rich Pereira – Deputy Director, Demography

What we will cover today

- The future vision for the Population and Migration Statistics System in England and Wales
- Progress update on building the research evidence base
- Next steps
- Privacy and disclosure control

Help Shape our Future 2018 White Paper

- “In 2023, ONS will present recommendations to government as to the future of census arrangements, in light of the progress that has been made. Consultation with, and the views of, end users and decision-makers will inform these recommendations in respect of how far the new system will be fit for purpose in meeting needs. The methodologies underpinning the new approach will also be quality assured by an expert external assurance panel*”

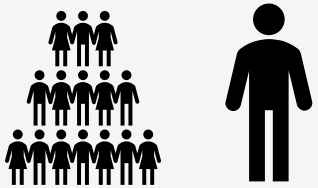
- *Methodological Assurance Review Panel

A reminder of our Current system – Mid Year Estimates

- Published in June, reference period is the previous year to mid point
- Experience 'drift' over the decade and adjusted post the next census
- Although they use a standard cohort component methodology there are necessary complexities that have grown organically over time and are not as transparent as we'd like them to be (for example, we don't share the code)
- Scope for more automation and production pipelines
- Financially complicated and 'programme heavy' (The Census is a stand-alone programme governed by special arrangement with our Treasury) - scope for smoothing finance profile
- Using this system can be challenging when meeting emerging user requirements such as increased frequency or different definitions

Future Population & Migration Statistics System

Strategic vision: More frequent, timely and inclusive statistics about the population and its characteristics



A social statistics system that shapes our understanding of society and people.

A system to create and share statistics that is sustainable and future proof.

A broad, flexible range of outputs and products to respond to user needs and questions of the day.

Administrative data will be at the core of this system.

The transformed outputs will include

National to local **Population and Migration Statistics**

Understanding demographic change and respond to challenges

Population sub-groups and characteristics

Reflecting the diversity of populations to local levels

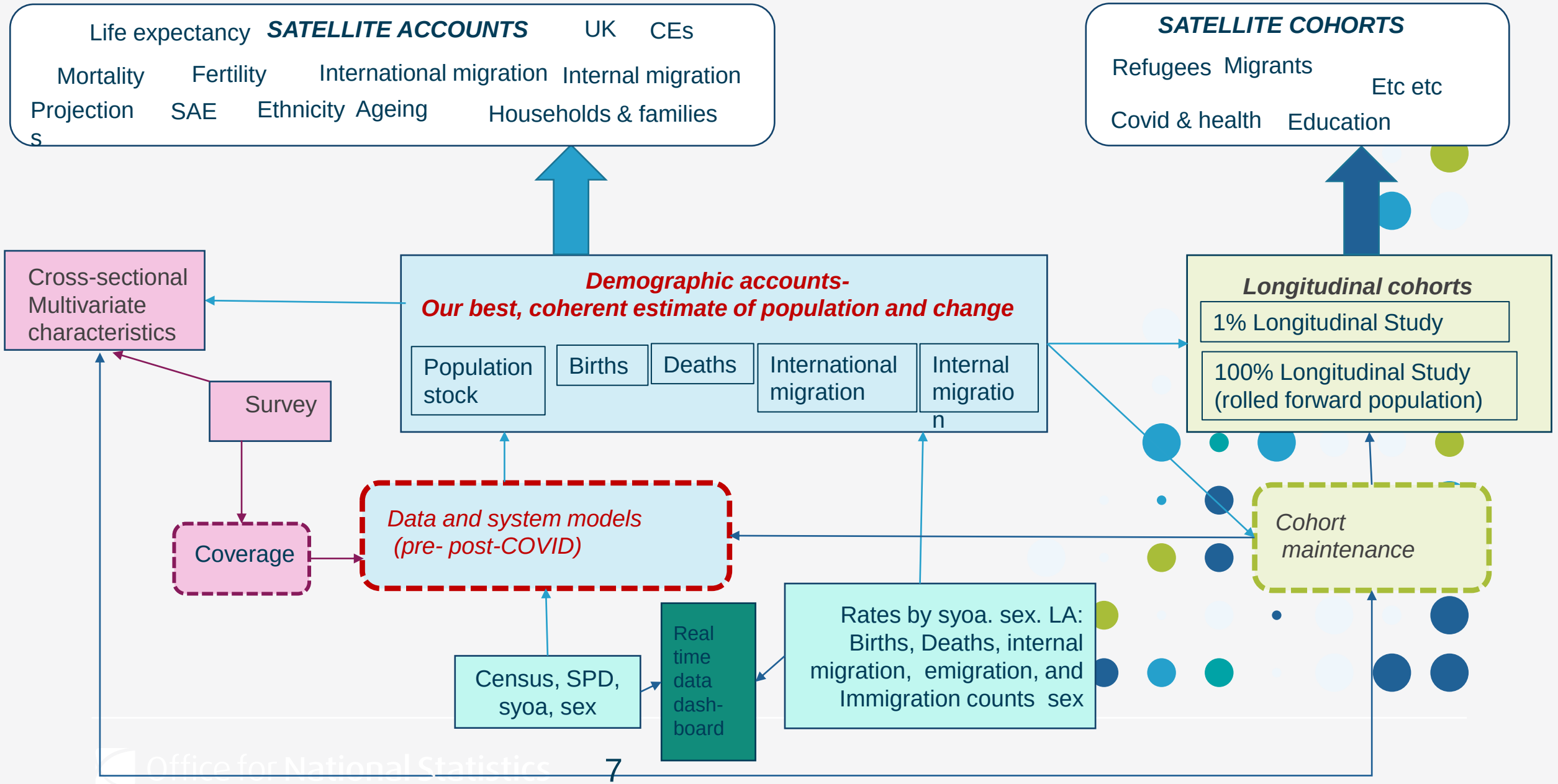
Housing, accommodation and living arrangements

Understanding who and how people live together in our society

Outcomes

Data assets that allow us to understand population life journeys and outcomes

Hybrid statistics system

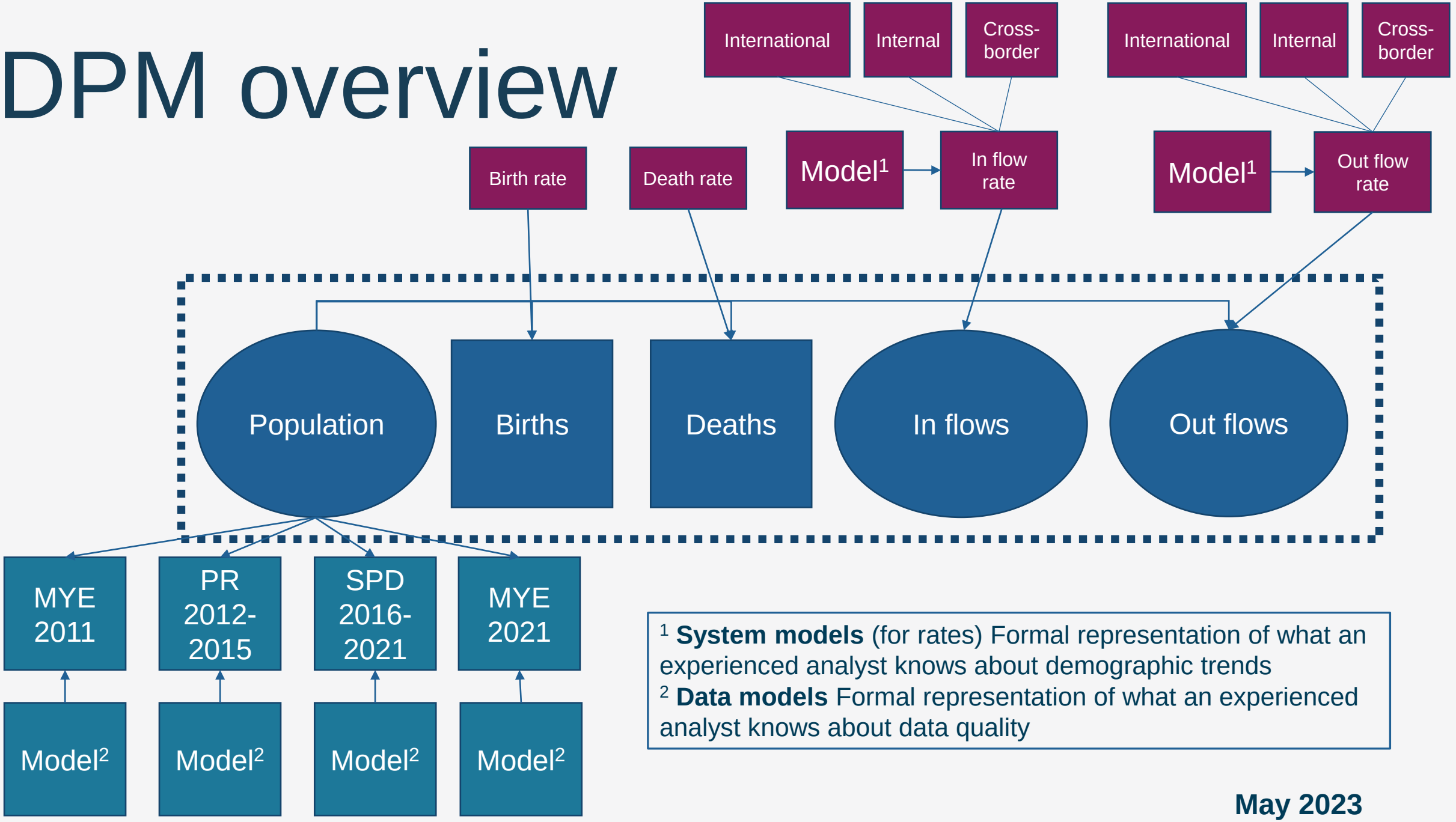


Transforming our population estimates – introducing the Dynamic Population Model (DPM)

What is the DPM?

- A statistical model to estimate the demographic account by combining data sources
- like the MYEs, it captures the dynamics of population change through the cohort component method of demographic accounting, ie adding births and immigrants, taking out deaths and emigrants.
- Is designed to benefit from additional data, uncertainty in data supply, and access to significantly more computing power
- Currently, the account is estimated annually from 2011 onwards by local authority, sex and single year of age
- The account is estimated separately for each year of age cohort
- We currently use Particle Filters for estimation (Sequential Monte Carlo method) - experimenting with Template Model Builder (TMB)

DPM overview



[1] **System models** (for rates) Formal representation of what an experienced analyst knows about demographic trends

[2] **Data models** Formal representation of what an experienced analyst knows about data quality

May 2023

Input data in the DPM

- Counts of births and deaths from registration data (assumed exact)
- Population stock data (point estimates, measures of uncertainty and coverage adjustment)
  - MYE 2011 (Census 2011-based)
  - PR 2011-2021
  - SPD 2011, 2016-2021
  - MYE 2021 (Census 2021-based)
- Smoothed superpopulation rates of births (by age of mother), deaths, in-migration and out-migration

Main steps for estimation

1. **Approximate components.** Build approximations of the series for births, deaths, migration, and population. Unlike in the final account, these series do not have to be mutually consistent.
2. **Fit system and data models.** Use the approximate series to fit models for births, deaths, and migration. The models all contain hyper-parameters, which are kept, and rates, which are discarded. Similar calculations, based on the approximate series for births, deaths, migration, and population, are also done for data models.
3. **Estimate individual accounts and rates.** Using the hyper-parameters and the raw data, (re-)estimate demographic accounts and rates. Each Local Authority is estimated independently (conditional on the hyper-parameters and data).
4. **Combine accounts, derive migration.** Combine the individual accounts into a unified account for all of England and Wales. As part of this process, derive values for all migration flows between Local Authorities, and between all Local Authorities and the outside world.

Admin Based Population Estimates (ABPEs)

- Using the Dynamic Population Model (DPM) we've produced ABPEs for 2011 to 2022 for all local authorities in England and Wales

- The results are showing promise and we've demonstrated the need for a coverage adjustment method which we plan to develop as part of next steps. The method might be like using a Census Coverage Survey as we do alongside our ten-yearly census

Different versions of the DPM (February 2023)

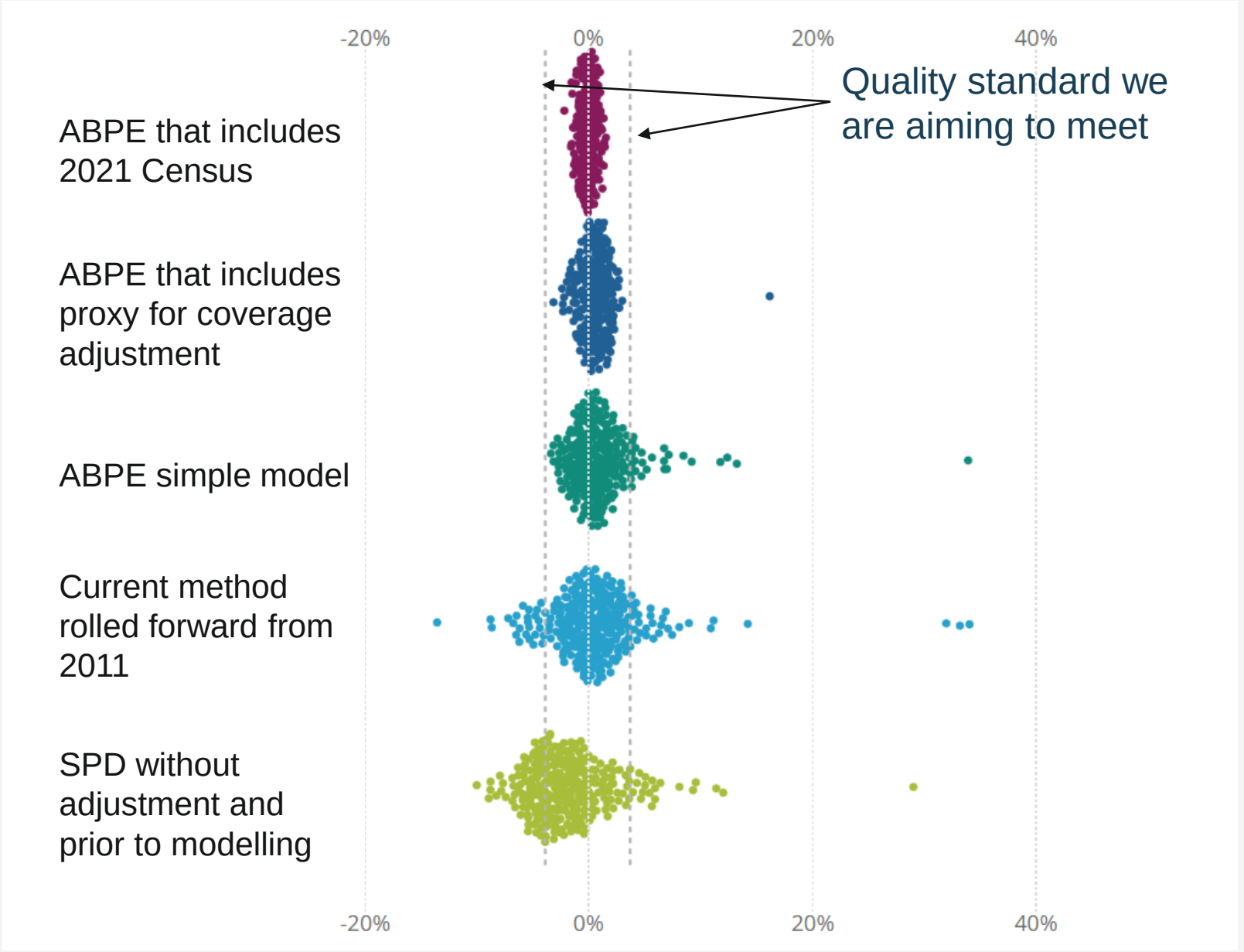
- ABPE Best Estimates
  - Uses MYE 2021 based on Census 2021 as a population stock
  - Uses interpolated coverage ratios between 2011-2021

- ABPE Future Estimates
  - Modified Best Estimates to replace MYE 2021 as a stock with SPD 2021

- ABPE Basic Estimates
  - Excludes Census 2021 entirely, comparable to MYE 2021 rolled forward from 2011
  - Uses 2011-based coverage ratios only

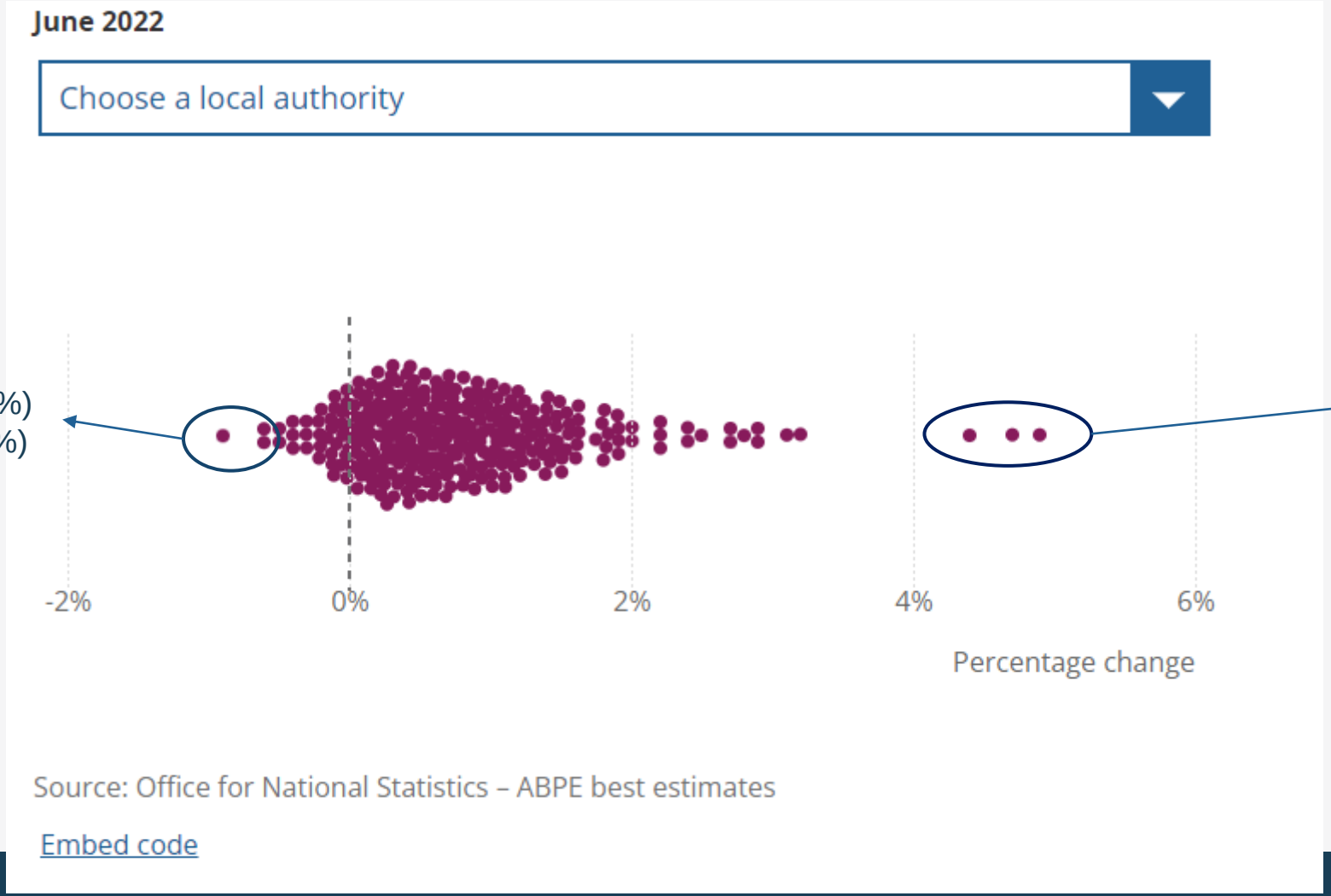
How do our admin based population estimates (ABPE) stack up at LA level?

- Results are showing potential
- Demonstrates the need for a robust coverage adjustment



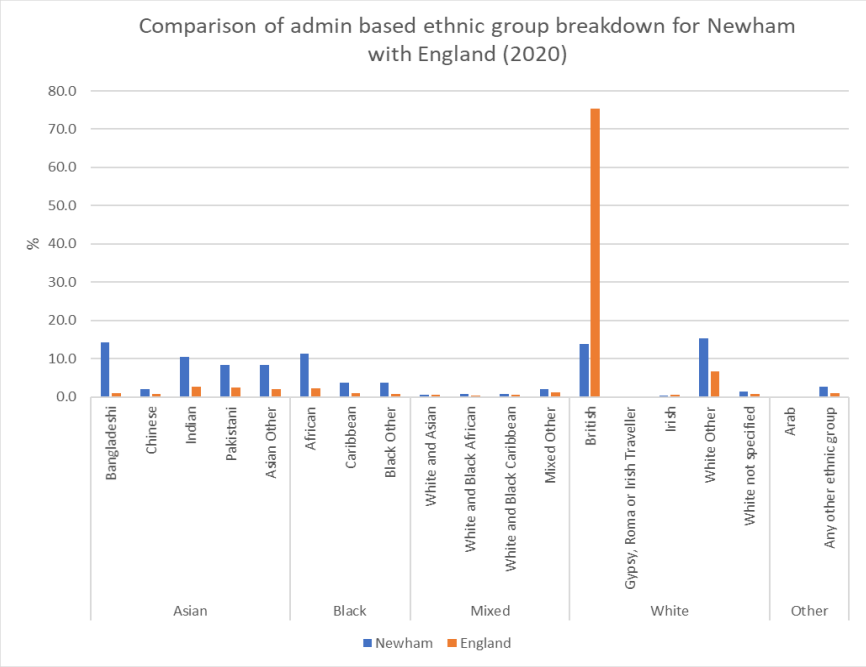
What do our ABPEs show about population change between 2021 and 2022? As expected, the largest population growth are seen in London local authorities. We have produced these estimates 4 months earlier than we do in the current system

From left to right:
- South Staffordshire (-0.9%)
- Kingston upon Hull (-0.6%)
- Lincoln (-0.6%)

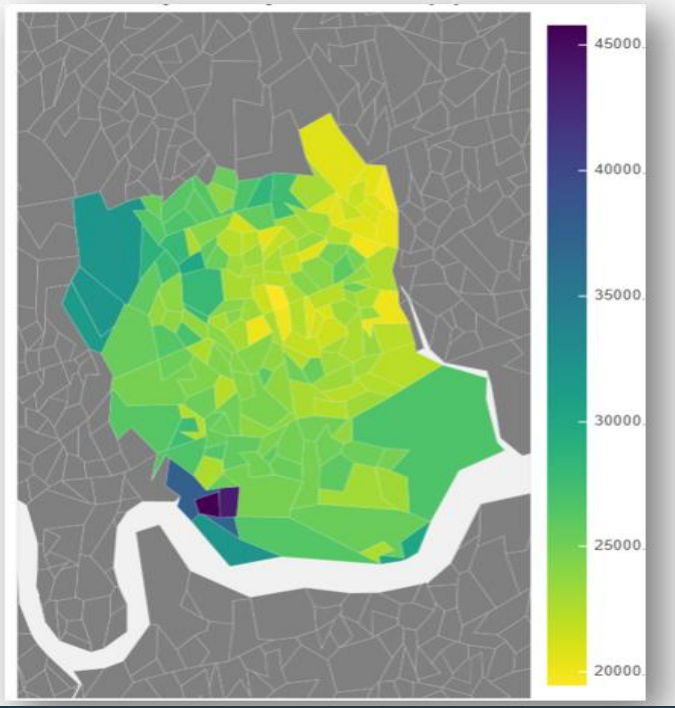


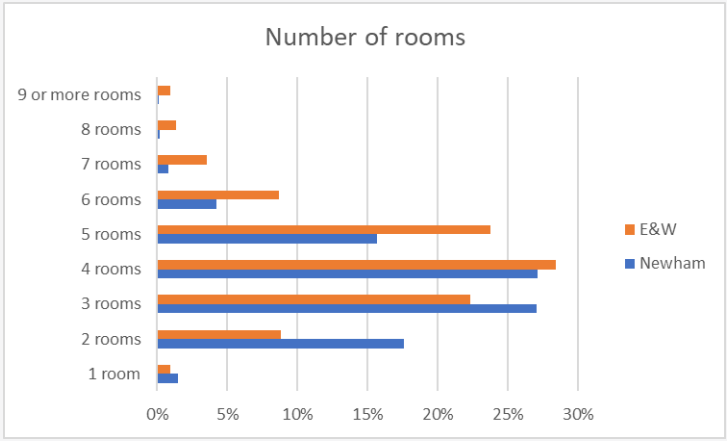
From left to right:
- Camden (+4.4%)
- Westminster (+4.7%)
- Tower Hamlets (+4.9%)

How is our characteristics work shaping up? Illustrated through a case study of Newham

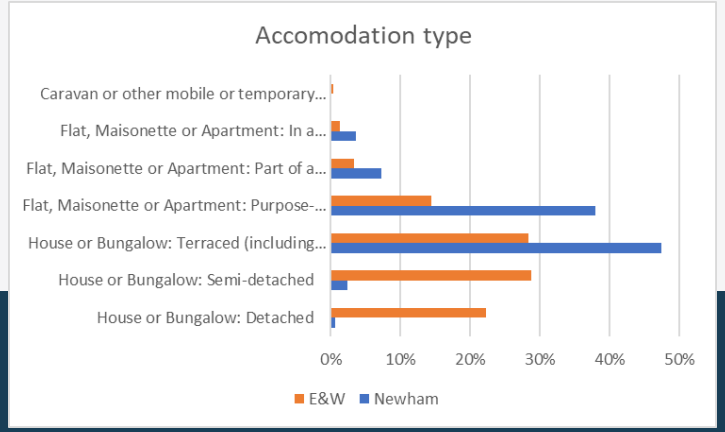


Median household income is lower in Newham than in England. At LSOA level there is more variation, with some affluent areas in the west (Canning Town, Stratford)





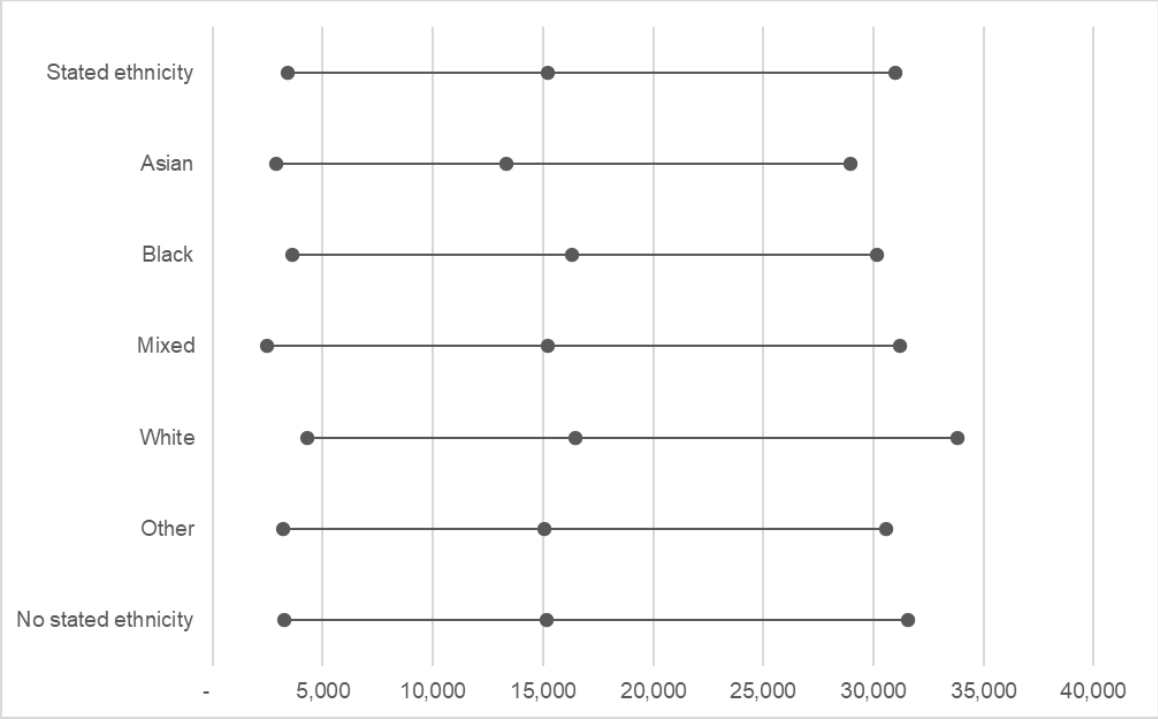
Housing stock in Newham (blue) typically has fewer rooms and more flats and terraced properties compared to England (orange)

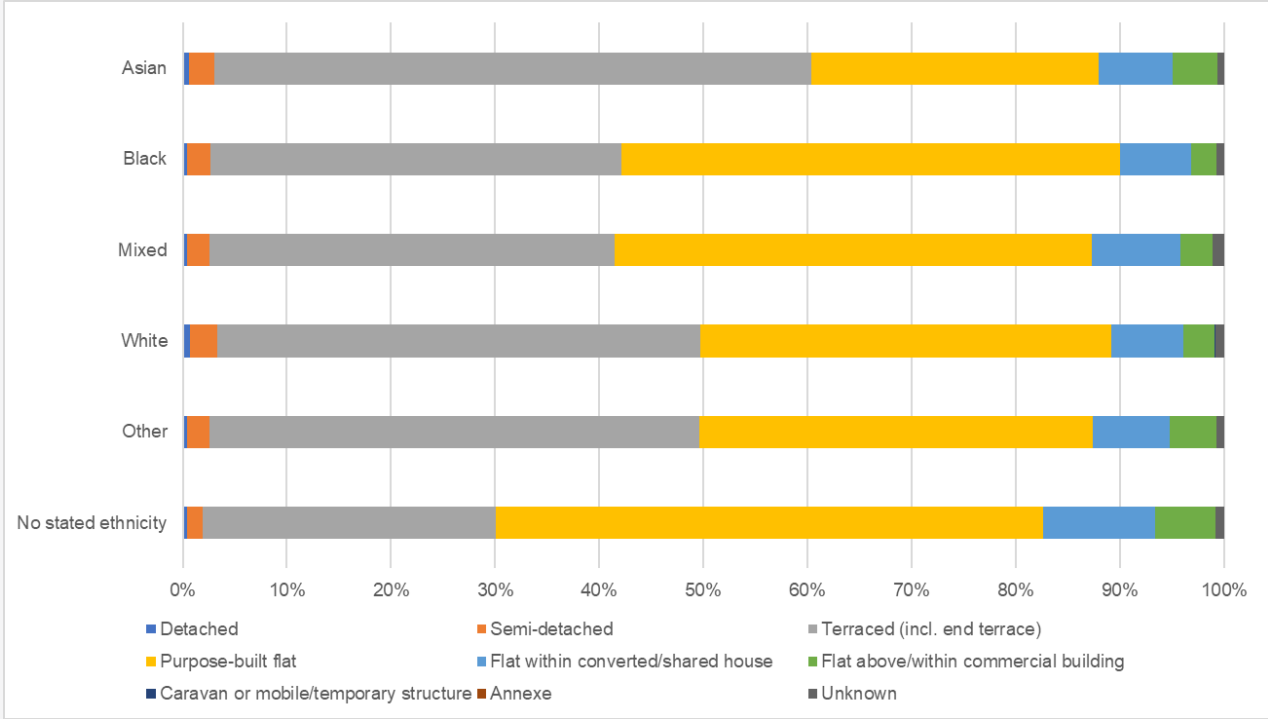


Newham (blue) has a more heterogeneous, non-white population compared to England (orange)

Staying with Newham, what do our multivariate findings show?

Individual median income by ethnic group for Newham shows very similar patterns seen for England overall





There is some variation seen in the most common housing type for different ethnic groups in Newham. Flats are the most common housing type for those with Black or Mixed ethnicity, while terraced properties are more common for those with White, Asian or Other ethnicity.

Operationalising the new system

- Data Quality - measuring and assuring
- Sustainable platform for Development, Testing and Production
- Sustainable data flows
- Transparency and reproducibility
- Efficiency in terms of data use and manual input
- Official Statistics badging process

Evidence we will deliver, user benefits

Improving existing outputs

Evidence:

Admin based population/migration estimates

Admin based housing and ethnicity statistics (small area univariate & multivariate)

Admin based occupied address (household) estimates

Meeting unmet user needs

Evidence:

Small area multivariate admin based income statistics

Pop24/7 daytime population base proof of concept

Admin based morbidity research

Future system

Improving scope of longitudinal outputs

Evidence:

Refugee Integration Outcomes study

Longitudinal Outcome of a Covid Infection

Updated Veteran estimates

Frameworks

Evidence:

Conceptual frameworks for Population and Migration Statistics

Quality measures

Quality framework for characteristics (including summary of methods for improving quality gaps)

Acceptance criteria (to transition from research to official statistics)

Other topics: further development

Admin data only ⟷ **Admin and survey data** ⟷ **Survey data only**

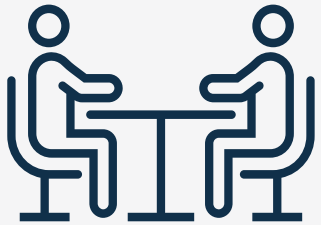
Annual publications at LSOA level	Annual publications at LA level, with development plans for LSOA	Less frequent OR less timely publications at LA level	Further research required
• Ethnicity • Income • Housing characteristics • Education • General Health • Disability	• Labour market status • Vehicle Ownership • Veterans • Household composition • Tenure	• Sexual Orientation • Main language • Country of Birth	• National Identity • Occupation • Religion • Caring responsibilities • Welsh Language • NS-SEC

We will publish a paper outlining this for users alongside the launch of the consultation. This will be accompanied with some methods that we might use to produce these outputs, which could be categorised as:

- Methods for producing outputs based on admin data alone
- Methods for producing outputs that combine survey and admin data
- Methods for producing small area outputs that are based on survey data alone

Consultation objective

- To understand how close our proposed system is to meeting user's needs
- To understand the future requirements of users to inform and prioritise our research plans



Next steps, alongside ongoing research

Public Consultation

- Hosted on consultations.ons.gov.uk
- Describing the transformed system, seeking views on progress, priorities and user needs
- Formal response published

Engagement

Including

- UK Government
- Devolved administrations
- Local and Combined authorities & other public bodies
- Business & charity sectors
- Public engagement
- Academia
- Parliament

Recommendations

- National Statistician's recommendations for England and Wales
- Progress on the transformation and recommendations on what is needed to deliver & support transformed statistical system

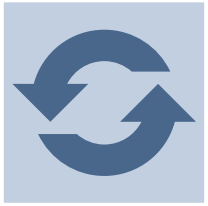
UK statistics and work with National Records of Scotland & NI Statistics and Research Agency

- Census responsibility is for England and Wales, and therefore the Recommendation is and we are working closely with Welsh Government on it
- NRS and NISRA developing their plans for future statistics system in parallel
- National Statistician, Registrars-General and Chief Statisticians signed agreement on co-operation in November 2022
- Ongoing engagement between all three agencies and with Welsh and Scottish Governments

Timeline

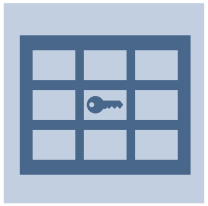

Consultation (Summer to Autumn)
Analysis of consultation responses (Autumn to Winter)
Summary of consultation responses and National Statistician's recommendation (Winter to Spring)
Publication of research plans and ongoing development (Spring onwards)

Disclosure Control – Current system



Targeted Record Swapping

Continued main source of protection



Cell Key Perturbation

Perturbation applied consistently

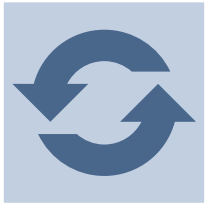
Protects against differencing

Automated Disclosure Checks

Rules to determine what is safe

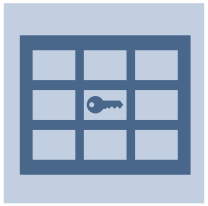
Much faster publishing

Disclosure Control – Future?



Targeted Record Swapping

Harder to implement in data that continuously gets updated and where records appear in many sources – we need an alternative



Cell Key Perturbation

Source data agnostic

Automated Disclosure Checks

Source data agnostic

Privacy measures to date:

- Segregation of duties – analysts don’t see personal identifiers such as name and full address
- RDMF – Reference Data Management Framework – data is linked using an ONS key that is generated in the RDMF
- Held in secure systems

Questions and comments welcome!