

Transformed integrated population and migration statistics: an update

Louisa Blackwell

Office for National Statistics

UNECE, September 2022

Contents

1. The wider context: ONS transformation
2. Limitations of the current system
3. The dynamic population model
4. The estimation window
5. Advantages and disadvantages
6. Pre-processing and the real-time data dashboard
7. Statistical population datasets and special populations
8. Admin-based migration estimates
9. ONS work on admin-based characteristics

Acknowledgement: Modelling Team in ONS and from University of Southampton: Peter Smith, Jakub Bijak, Jason Hilton. Also the Longitudinal Scientific Advisory Panel.

Why transform?

Aligning to **UKSA Strategy**, ‘Statistics for the Public Good’

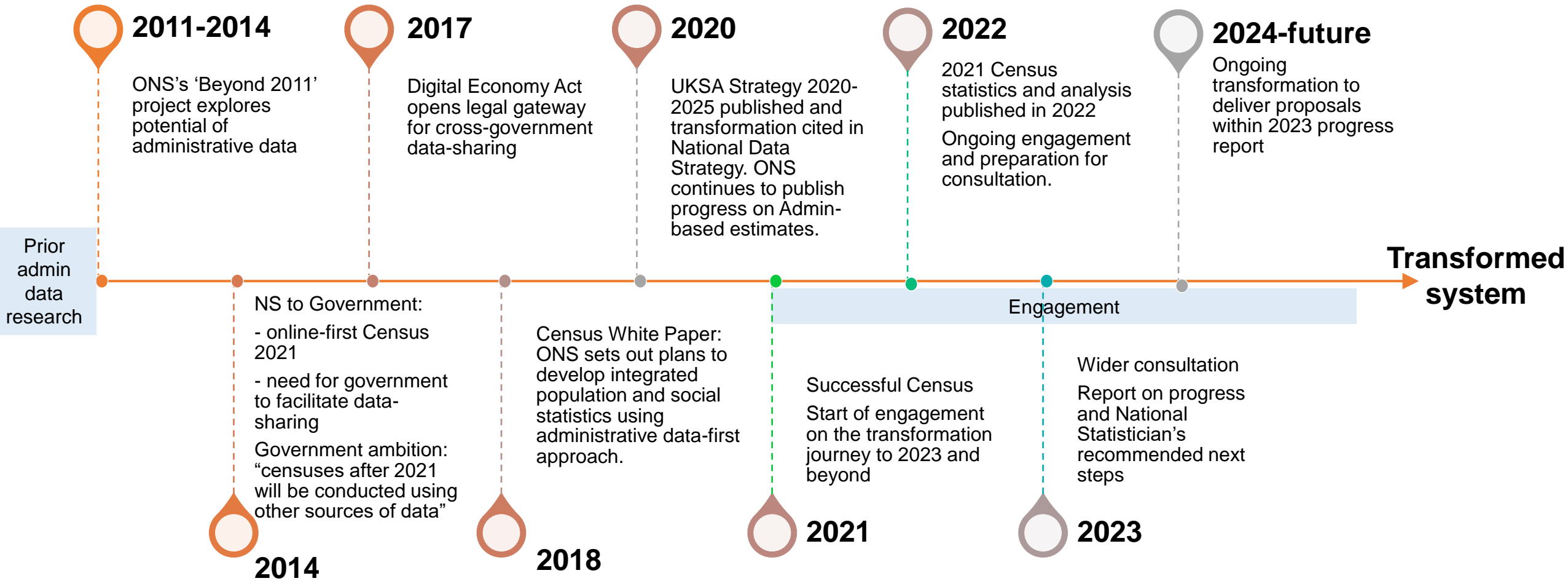
- **Sustainable:**
 - Using **best available information**,
 - Maintaining **consistent quality** over time;
- **Radical:**
 - Understanding need for more frequent estimates;
 - Considering **radical change** to system;
- **Ambitious:**
 - Aiming to produce **more frequent & timely estimates**.
- **Inclusive:**
 - Aiming to produce **more inclusive statistics**, over time.



*“The pandemic has **revolutionised** public data and **there’s no turning back**”*

- Sir Ian Diamond

Timeline



Our ambition

Provide the best picture of the UK's population in the long term, moving towards a population and migration system that will deliver

- Richer, more detailed and more timely updates to census information than we've ever had before
- Increases in breadth and detail over the coming years

Limitations of the existing population estimation system

1. Manual, with substantial expert judgment

- Slow
- Difficult to replicate

2. Inflexible

- Difficult to change data sources or outputs

3. **Critical dependence on Census**

- **Growing uncertainty and decennial rebasing**

Emerging challenges

- Demand for more frequent updates
- Demand for more granular output
- Possible loss of traditional census
- Maximise use of administrative data
- Changes to input data
- Appearance of new data sources, eg cell phones

Structure

Demographic account- John Bryant and Juanni Zhang

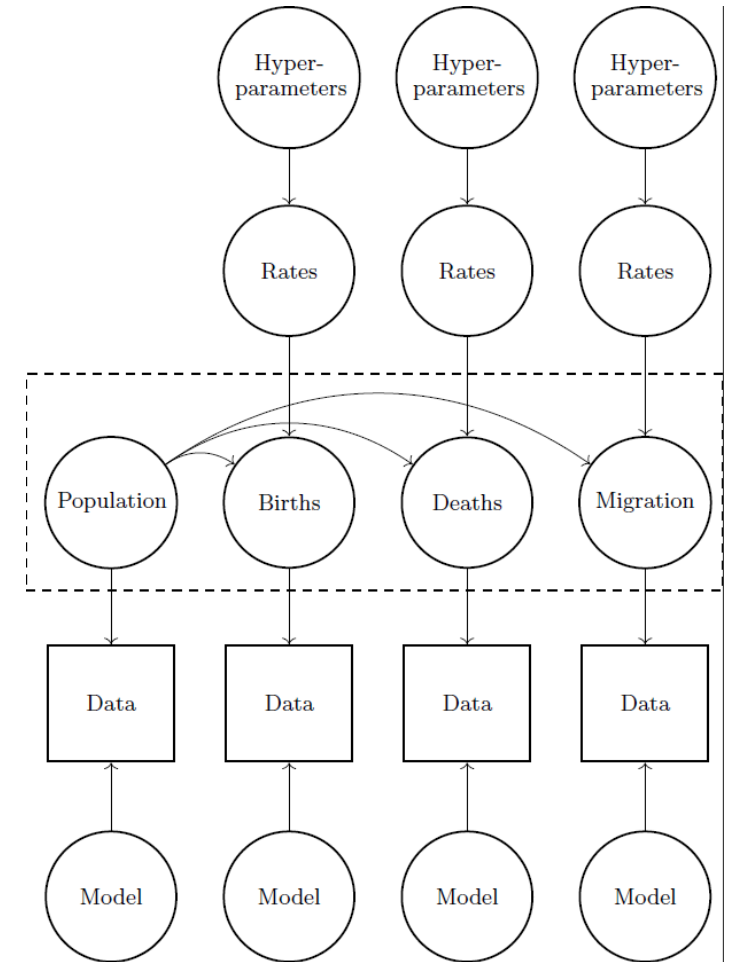
- Demographic equivalent of national account
- Standardised format
- Consistency between stocks and flows

System models

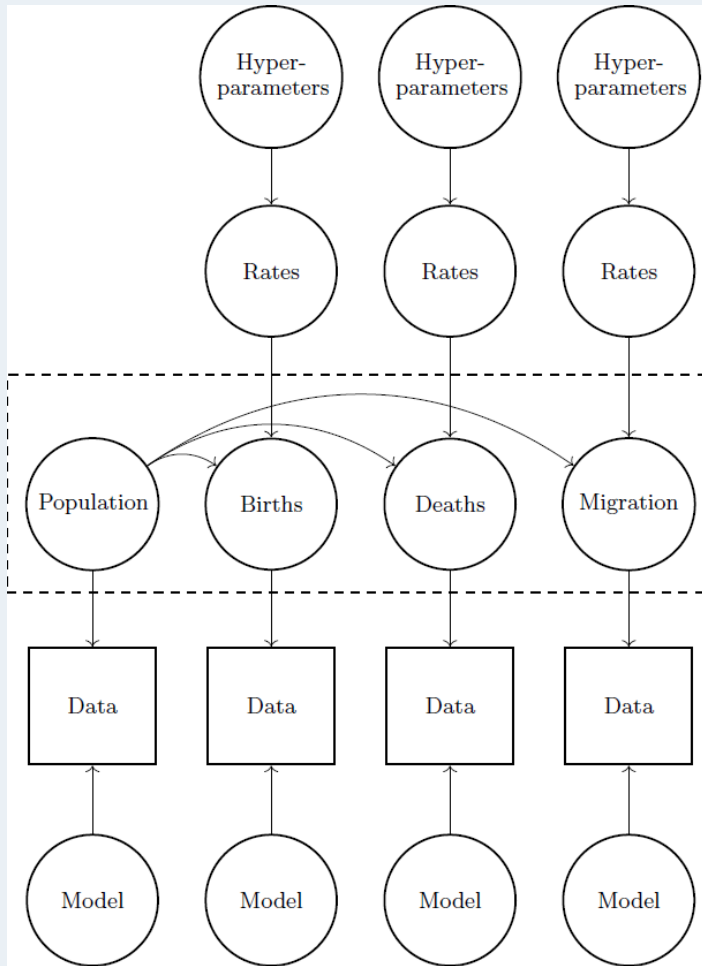
- Statistical models of regularities in birth, death, migration rates
- Formal representation of what an experienced analyst knows about demographic trends

Data models

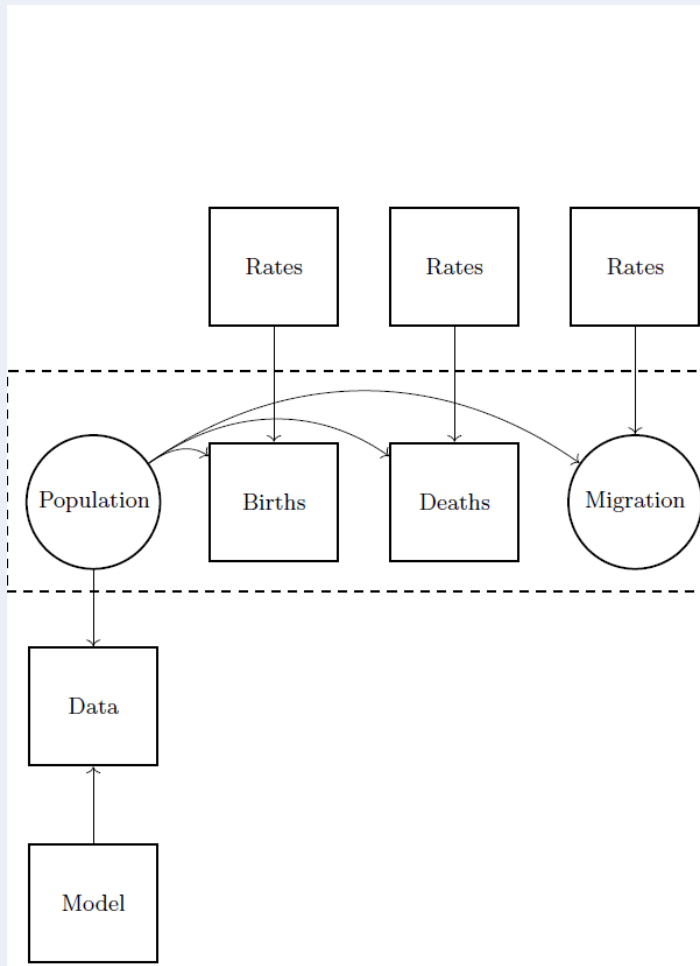
- Statistical models measurement, coverage errors
- Formal representation of what an experienced analyst knows about data quality



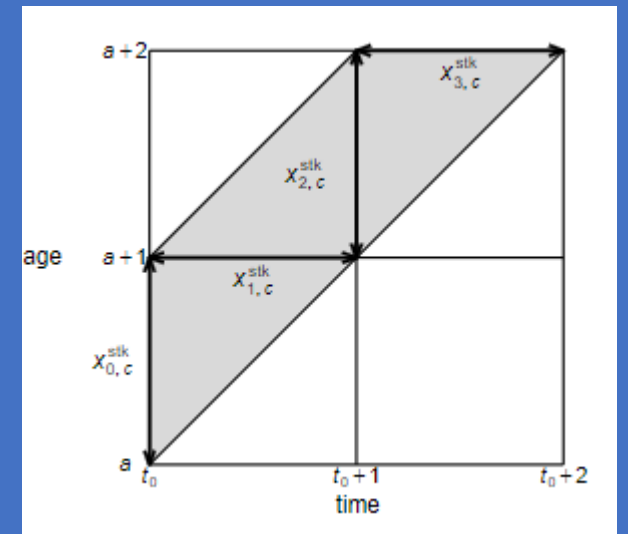
Full model



Simplified



Particle filters within cohorts



Estimation windows for Dynamic Population Model

Assumptions: 1) 15 month lag in availability of components (maximum lag from international migration?)
 2) Key predictors are available monthly

| Demographic accounts | | 2021 | | | | | | 2022 | | | | | | 2023 | | | | | | | | | | | | | | | |
|----------------------|--------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Estimates | | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun |
| Monthly Reporting | Jun-22 | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange |
| | Jul-22 | Blue | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange |
| | Aug-22 | Blue | Blue | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange |
| | Sep-22 | Blue | Blue | Blue | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange |
| | Oct-22 | Blue | Blue | Blue | Blue | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | Orange | Orange | Orange | Orange | Orange | Orange | Orange |
| | Nov-22 | Blue | Blue | Blue | Blue | Blue | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | Orange | Orange | Orange | Orange | Orange | Orange |
| | Dec-22 | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | Orange | Orange | Orange | Orange | Orange |
| | Jan-23 | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | Orange | Orange | Orange |
| | Feb-23 | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | Orange | Orange |
| | Mar-23 | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | Orange |
| Apr-23 | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Grey | Orange | |

Finalised estimates from DPM using finalised predictors

Best monthly provisional estimates from DPM based on incomplete predictors

Forecasts based on historic time series and expert opinion

Advantages

Automation

- *Permits frequent updating, detailed outputs*

Transparency

- *Replicable*
- *Permits experimentation*

Flexibility

- *Easy to change inputs, outputs*

Uncertainty

- *Formal measures of uncertainty*

Timeliness

- *Uses early aggregate-level data*

Disadvantages

Computation

Like much Bayesian computation, current methods are slow
Current software not scalable to 330 Local Authorities

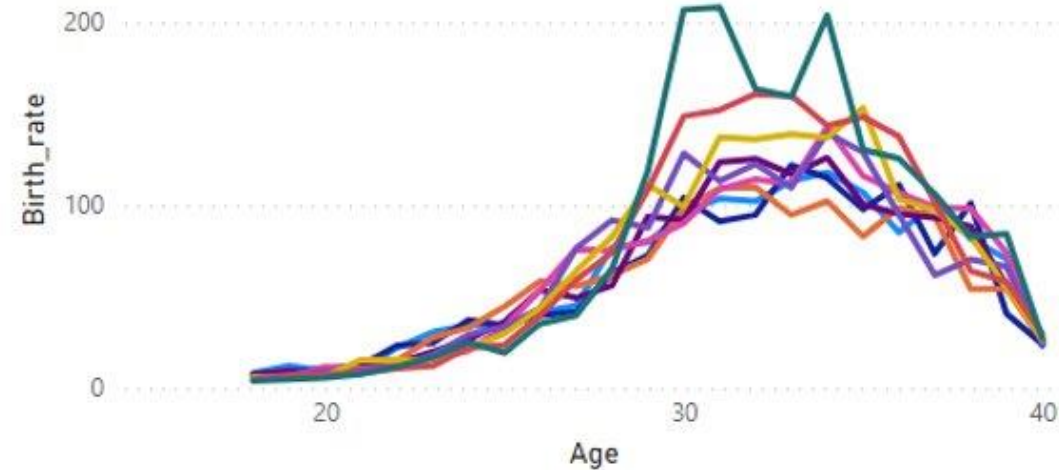
Highly technical

Need to think about communication with users
Requires specialist statistical computing skills at ONS

Data development

Birth_rate by Mother's Age and Year

Year ● 2012 ● 2013 ● 2014 ● 2015 ● 2016 ● 2017 ● 2018 ● 2019 ● 2020



- Real-time data dashboard
 - Input checks
 - Output checks
 - Calibration of forecasts
- Formal statistical framework exposes data issues
- Involving demographic experts in assessment of input data

Model performance

We are testing DPM performance through simulation studies

This involves simulating the demographic account (population stocks and flows) for 2011-2020 to consider as a 'gold standard truth' that we are trying to estimate

We then test how the DPM performs under different scenarios

Scenarios include:

- putting in and taking out unbiased population stock data
- combinations of different types of error in the flow rates

So far we have found:

- The DPM produces more accurate estimates if we include unbiased population stocks in the estimation
 - For instance, if our migration estimates are biased then the DPM with population stock data applies a better correction than if the stock data are left out
 - If immigration and emigration are both either underestimated or both over-estimated, the marginal advantage of including stocks data in the estimation is small
 - If one is over-estimated and the other under-estimated, then the marginal advantage is more substantial

Example: females in a university town

When immigration and emigration are both over-estimated by 5%:

| | Relative bias | |
|-------------|---------------|------------|
| | without stock | with stock |
| Immigration | 5.0% | 4.8% |
| Emigration | 5.1% | 4.8% |
| Population | 0.04% | 0.02% |

When immigration is over-estimated and emigration is under-estimated by 5%:

| | Relative bias | |
|-------------|---------------|------------|
| | without stock | with stock |
| Immigration | 4.8% | 1.4% |
| Emigration | -3.5% | 1.4% |
| Population | 1.30% | 0.02% |

2012, 100 runs

Model development

- Moving to endogenous rates
- Development of data models
 - Input from demographic experts
 - DI/Census linkage
 - Uncertainty estimates
- Special population adjustments
- Parallelisation

Publications

July 2022

Synthetic Local Authority

September 2022

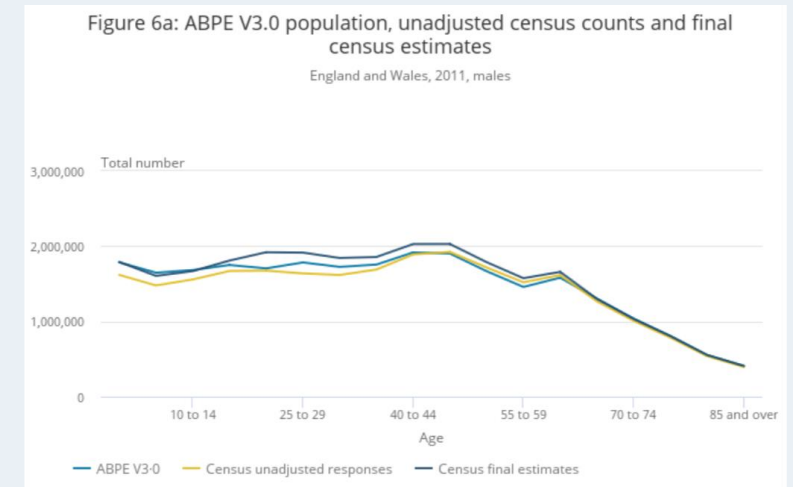
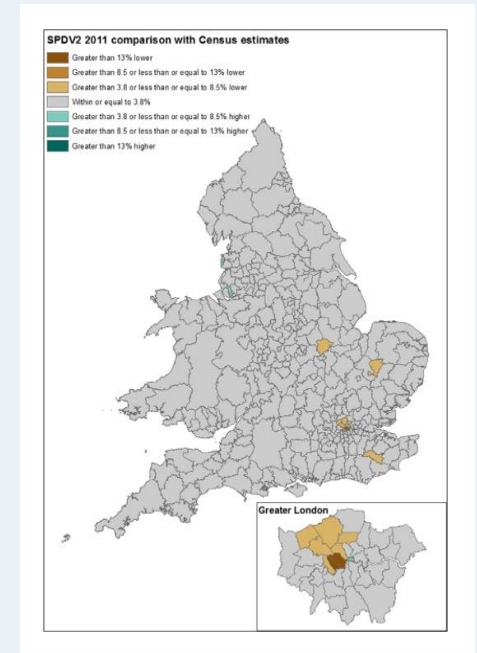
14 LA case studies 2022

December 2022

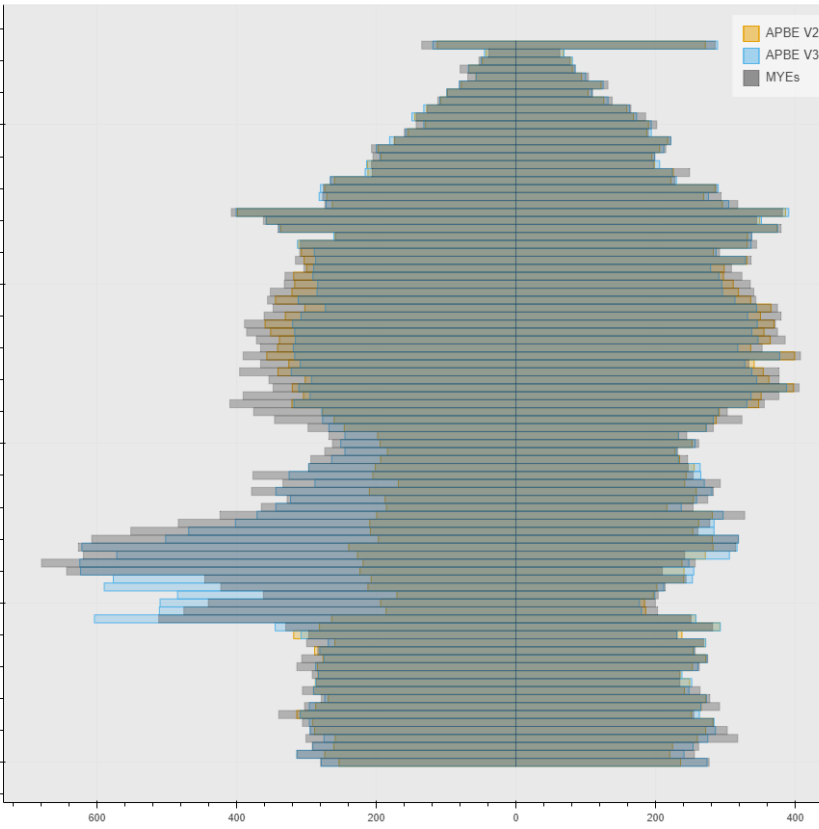
All LAs 2022

Transformed population statistics: Latest progress to date

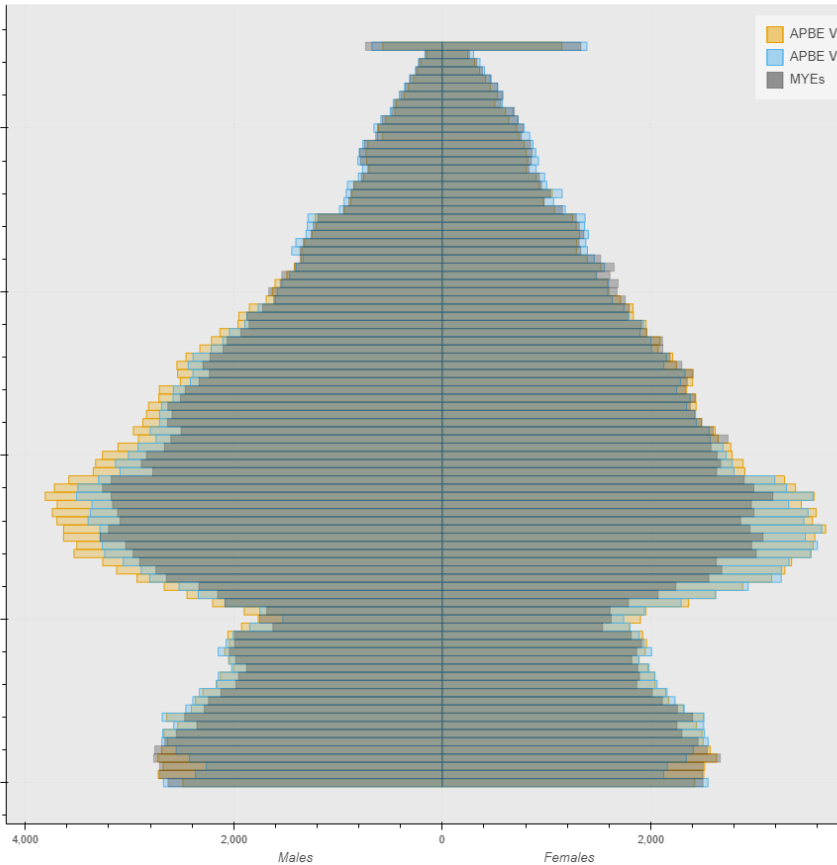
- **Population estimates**
 - Statistical Population Datasets (SPDs) created by applying rules to linked data sources to optimise coverage (i.e. accuracy)
 - Our SPDs been used during the Census quality assurance work and have received positive feedback from local authorities



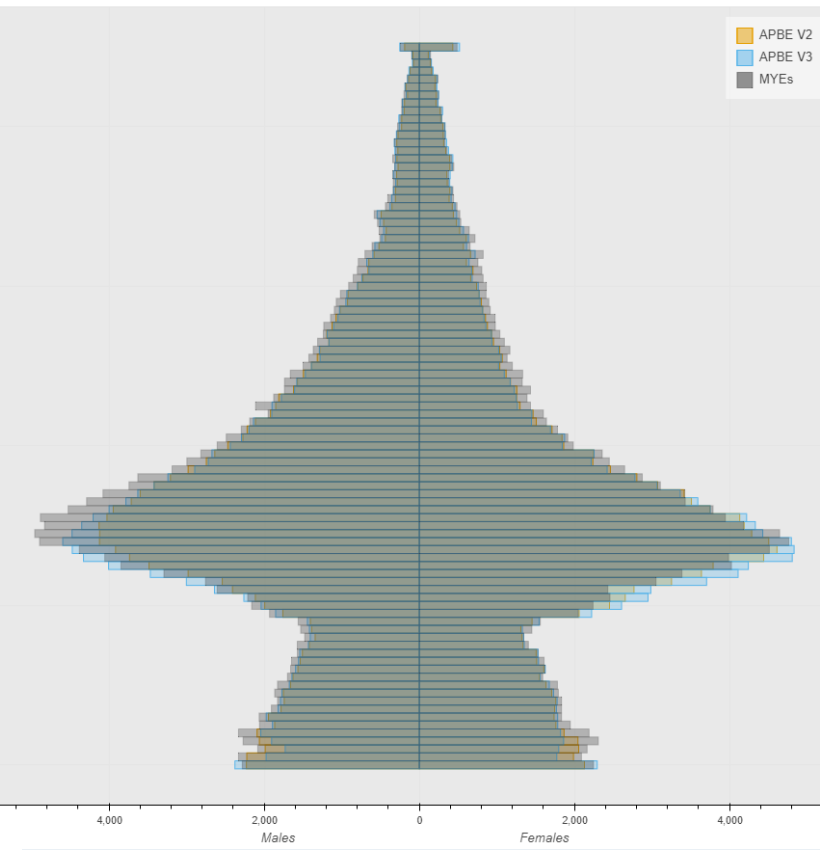
Richmondshire mid-2016

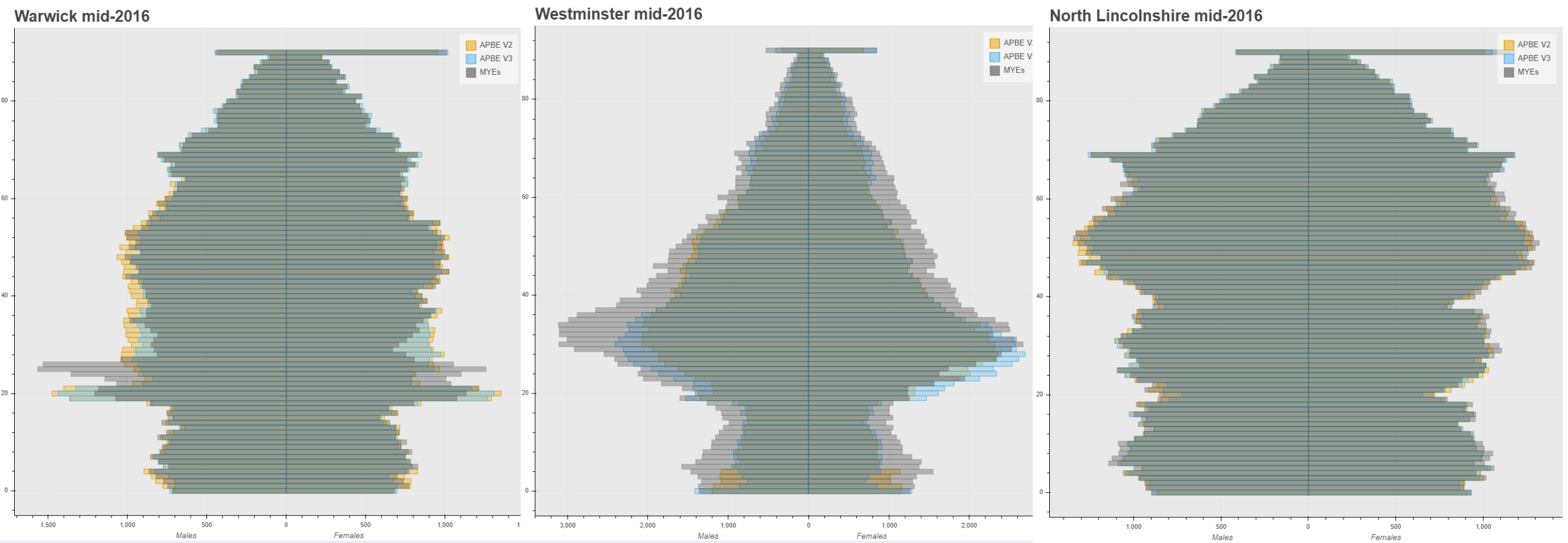


Ealing mid-2016



Tower Hamlets mid-2016





Grey bars – official Mid-Year Estimates
 Yellow bars – ABPE v2
 Blue bars – ABPE v3

Conclusion: Not a “one size fits all” approach

Next steps – using what we’ve learned to develop SPDv4

Historic approach to migration estimates

International Passenger Survey (IPS)

Estimates based on migrants' intentions



Northern Ireland migration flows (NISRA)

Asylum seekers (Home Office)

Adjustment for people changing their intentions



Long-term International Migration (LTIM)



Migration Statistics Quarterly Report: August 2020

A summary of the latest long-term international migration estimates for the UK for the year ending March 2020. Data from the Home Office, Department for Work and Pensions (DWP) and Higher Education Statistics Agency (HESA) are also included.

This is the latest release. [View previous releases](#)

Correction

Contact:
Mike James

Release date:
27 August 2020

Next release:
To be announced

Table of contents

1. Other migration outputs in this release
2. Development of migration statistics
3. Main points
4. Migration to and from the UK
5. EU and non-EU migration over time
8. Insights on recent international travel patterns
9. International migration data
10. Glossary
11. Measuring the data

Print this statistical bulletin

Download as PDF

Admin-based migration statistics – May 2022

Non-EU = Home Office visa data



EU = RAPID (DWP taxes and benefits)



British nationals = Previous (IPS-based) Modelled estimates



Immigration Emigration Net Migration EU, Non-EU and GB



Department for Work & Pensions

Achieving our vision for transformed migration and population statistics

“To use the best information to produce more frequent, timely and inclusive statistics that are coherent and flexible to evolving user needs”

Reinforcing
methods
with new
sources

Timeliness

Coherence

Reason for
migration

Alternative
definitions

Social Statistics- Admin First

- **Aim:** Transform the way ONS produces social statistics across a range of core topics (population/household characteristics) at the level of detail and geographical disaggregation required by users
- **Priority for 2022:** Deliver two admin-based multivariate case studies as evidence towards the 2023 Recommendation, covering:
 - Income by ethnicity
 - Housing by ethnicity

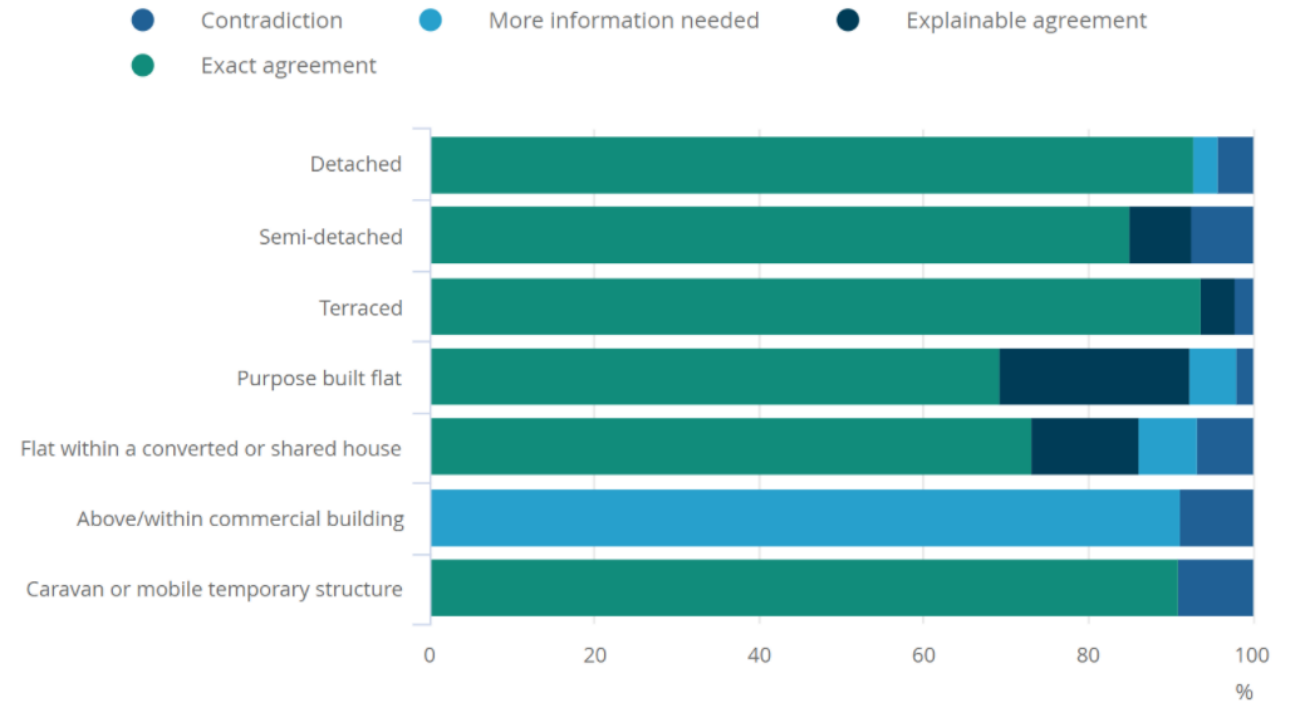
Admin-based successes: Housing stock

Admin-based housing stock characteristics

Property type, number of (bed)rooms, floor area

- Exploring the use of administrative data on housing as a replacement for collecting this information in censuses and surveys.
- Comparison of property type data recorded by the Valuation Office Agency (VOA) with data collected by the 2011 Census in England and Wales.
- Exact agreement between the VOA property type variable and the Census accommodation type variable for 86% of linked addresses

Agreement between census accommodation type and VOA property type



Source: Admin-based statistics for property type, feasibility research: England and Wales

Office for National Statistics

Admin-based successes: Ethnicity, Income, Labour Market

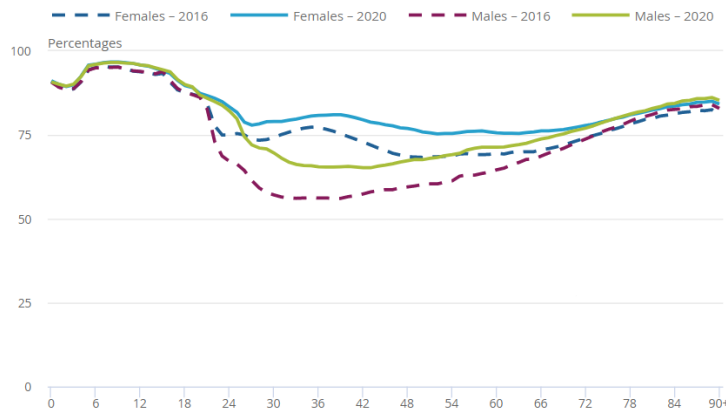
Admin-based ethnicity statistics (ABES)

Ethnicity information from HES, IAPT, ECDS, HESA, Birth Notifications linked to the 2016 ABPE V3 and admin-based ethnicity statistics produced.

Able to establish an ethnicity for 74.7% of individuals in the 2016 admin-based population base.

Figure 2: The proportion of people with a stated ethnicity increased for all ages between 2016 and 2020

Proportion of people in the ABPE with a stated ethnicity by age and sex, 2016 and 2020 admin-based ethnicity statistics, England



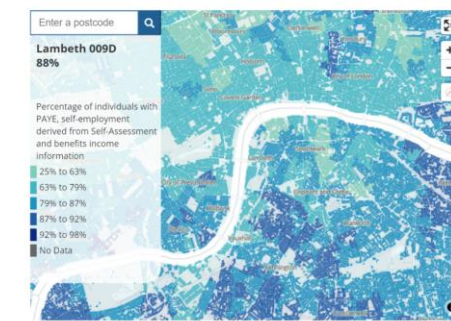
Admin based income statistics (ABIS)

Combine HMRC + DWP data, link to admin-based population and household estimates.

Individual and household net income estimates at LSOA (lower than ONS' survey-based outputs).

98.4% of occupied addresses (households) have income information from one source.

Percentage of individuals with PAYE, self-employment derived from Self-Assessment and benefits income information



Source: Admin-based income, England and Wales: tax year ending 2016 revised results. Office for National Statistics

Admin based labour market statistics (ABLMS)

Coverage is high, with 95% of individuals aged 16 years and over (on our population base) being assigned an admin-based labour market status (ABLMS).

Figure 5: Percentage of individuals assigned an ABLMS remains at or above 90% from age 20

Percentage of individuals with an assigned ABLMS, England and Wales, tax year ending 2016



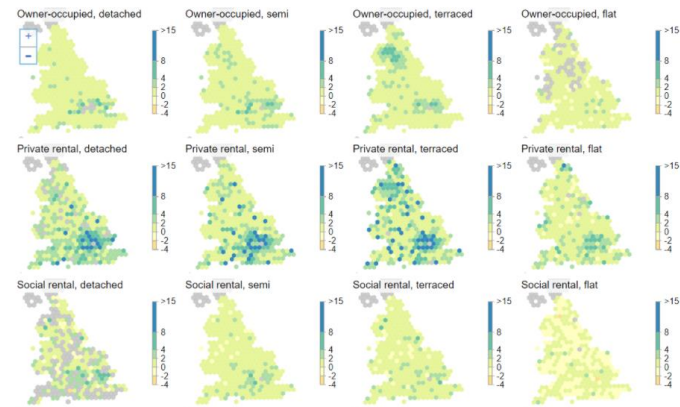
Admin-based successes: Overcrowding, Qualifications

Admin-based levels of overcrowding feasibility research

Assessed impact of replacing number of rooms question on Census 2021 by comparing sub-regional levels of overcrowding

Percentage point (pp) differences in the levels of overcrowding at LA level for 2011 Census number of bedrooms and VOA number of bedrooms using the bedroom standard by tenure and accommodation type, England & Wales

Households are identified as overcrowded if they have a bedroom occupancy rating of “-1 or less” using the bedrooms standard.



Source: Admin-based levels of overcrowding (using the bedroom standard and Valuation Office Agency number of bedrooms), feasibility research: England and Wales: January 2021

Office for National Statistics

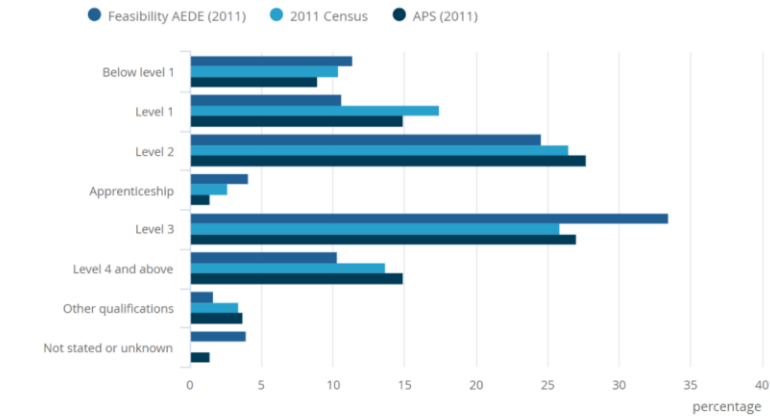
Admin-based highest level of qualifications feasibility research

Comparing admin and 2011 Census data, highest level of was the same on both sources for 57% of people.

For 84% of people, highest qualification level from admin data either agreed with, or was within one level of that recorded by the census.

Administrative data recorded a lower percentage of individuals with “Level 1” and “Level 4 and above” qualifications but a higher percentage with “Level 3”

Highest level of qualification in 2011 for usual residents aged 16 to 24 years, from the feasibility AEDE, 2011 Census and APS (percent distribution), England



Source: Admin-based qualification statistics, feasibility research: England

Office for National Statistics