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

Office for National Statistics
ONS’s ‘Beyond 2011’ project explores potential of administrative data

2011-2014

Digital Economy Act opens legal gateway for cross-government data-sharing

2017

2020


Government ambition: “censuses after 2021 will be conducted using other sources of data”

2014

Successful Census Start of engagement on the transformation journey to 2023 and beyond

2018

2021

Census White Paper: ONS sets out plans to develop integrated population and social statistics using administrative data-first approach.

2022

2024-future

2021 Census statistics and analysis published in 2022
Ongoing engagement and preparation for consultation.

Wider consultation Report on progress and National Statistician’s recommended next steps

2023

2024-2025 published

Transformed system

Prior admin data research
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

Office for National Statistics
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

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

- 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%:

<table>
<thead>
<tr>
<th></th>
<th>Relative bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>without stock</td>
</tr>
<tr>
<td>Immigration</td>
<td>5.0%</td>
</tr>
<tr>
<td>Emigration</td>
<td>5.1%</td>
</tr>
<tr>
<td>Population</td>
<td>0.04%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Relative bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>without stock</td>
</tr>
<tr>
<td>Immigration</td>
<td>4.8%</td>
</tr>
<tr>
<td>Emigration</td>
<td>-3.5%</td>
</tr>
<tr>
<td>Population</td>
<td>1.30%</td>
</tr>
</tbody>
</table>

2012, 100 runs
<table>
<thead>
<tr>
<th>Model development</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving to endogenous rates</td>
<td>July 2022</td>
</tr>
<tr>
<td>Development of data models</td>
<td>Synthetic Local Authority</td>
</tr>
<tr>
<td>• Input from demographic experts</td>
<td>September 2022</td>
</tr>
<tr>
<td>• DI/Census linkage</td>
<td>14 LA case studies 2022</td>
</tr>
<tr>
<td>• Uncertainty estimates</td>
<td>December 2022</td>
</tr>
<tr>
<td>Special population adjustments</td>
<td>All LAs 2022</td>
</tr>
<tr>
<td>Parallelisation</td>
<td></td>
</tr>
</tbody>
</table>
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
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)

Adjustment for people changing their intentions

Asylum seekers (Home Office)

Long-term International Migration (LTIM)
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
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”
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
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.

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.

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).
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

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.