

Scotland's Census 2021 Use of Administrative Data

Dr. Andrew Waugh

PAMS: 20 September 2018

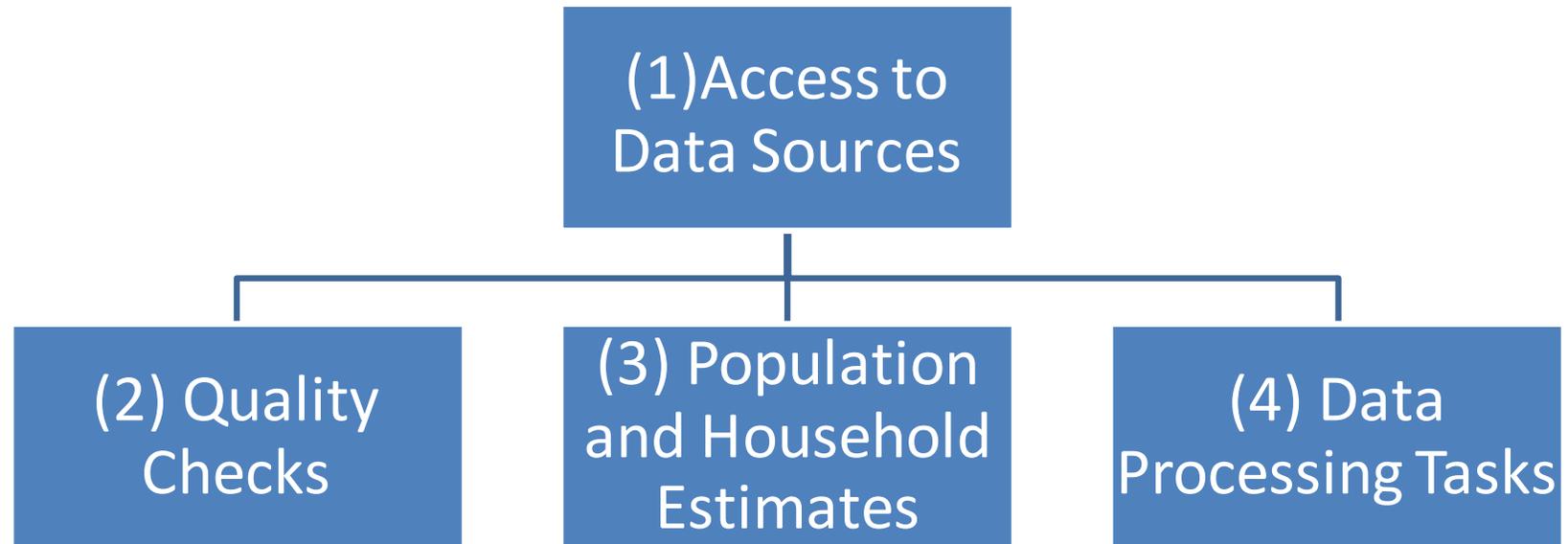
Contents

1. Why?
2. Objectives for the project
3. What Administrative Data do we have access to?
4. Why not use a single source?
5. Data Quality
6. Population and Household Estimates
7. Data Processing Tasks
8. Summary

Why?

- National Records of Scotland (NRS) should plan to carry out a census in 2021 which is predominately online but which should **make the best use of admin data** where possible.
- The scope of this census includes continuing to **explore the future greater use of admin data** in collecting socio-demographic statistics.
- There may be an opportunity to provide **enhanced outputs though the use of admin data** and this will be explored.
- **Make recommendations** for future censuses in Scotland.

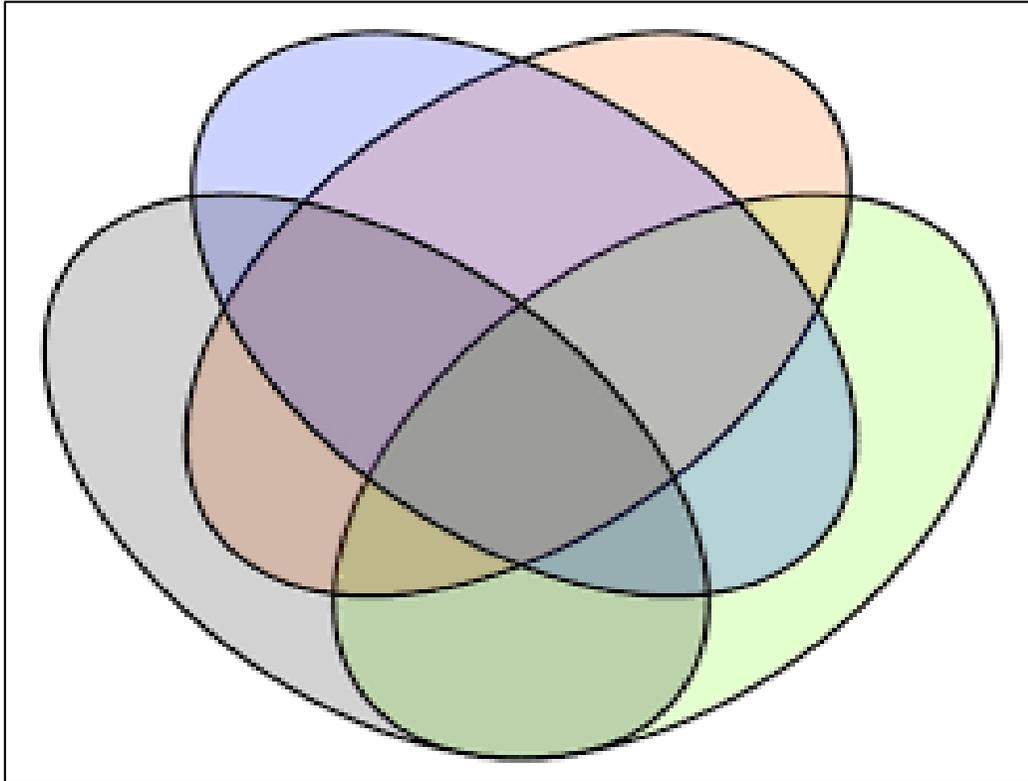
Objectives for the Project



(1) What Administrative Data do we have access to?

- NHS Central Register
 - Vital Events – Births, Deaths, Marriages & Civil Partnerships
 - Health Activity Data (contains no medical information!)
 - Electoral Registers
 - Higher Education Student Data (HESA)
 - Scottish Government School Pupil Census
 - Scottish Funding Council Further Education Student Data
 - Registers of Scotland (RoS) Residential Sales
 - NRS Geography Data
- Typically first name, last name, date of birth, gender, address, postcode. Very limited additional information.

Why not use a single source?



- No single source has coverage of the entire population
- Need to identify overlaps so we don't count people twice.
- Adjust for under-coverage.

(2) Data Quality

- Is the data we're using correct?
- Produce Quality Assessment of Admin Data Reports for each source.
 - E,g, [Report on internal migration for Scotland, quality assurance of administrative data used in population and migration statistics: January 2018](#)



Report on internal migration for Scotland, quality assurance of administrative data used in population and migration statistics: January 2018

Quality assurance undertaken on administrative data for internal migration for Scotland data used within Population Statistics Division (PSD) and Migration Statistics Division (MSD) publications.

Contact: Pete Large pop.info@ons.gov.uk +44 (0)1329 444661	Release date: 18 January 2018	Next release: To be announced
---	----------------------------------	----------------------------------

Table of contents

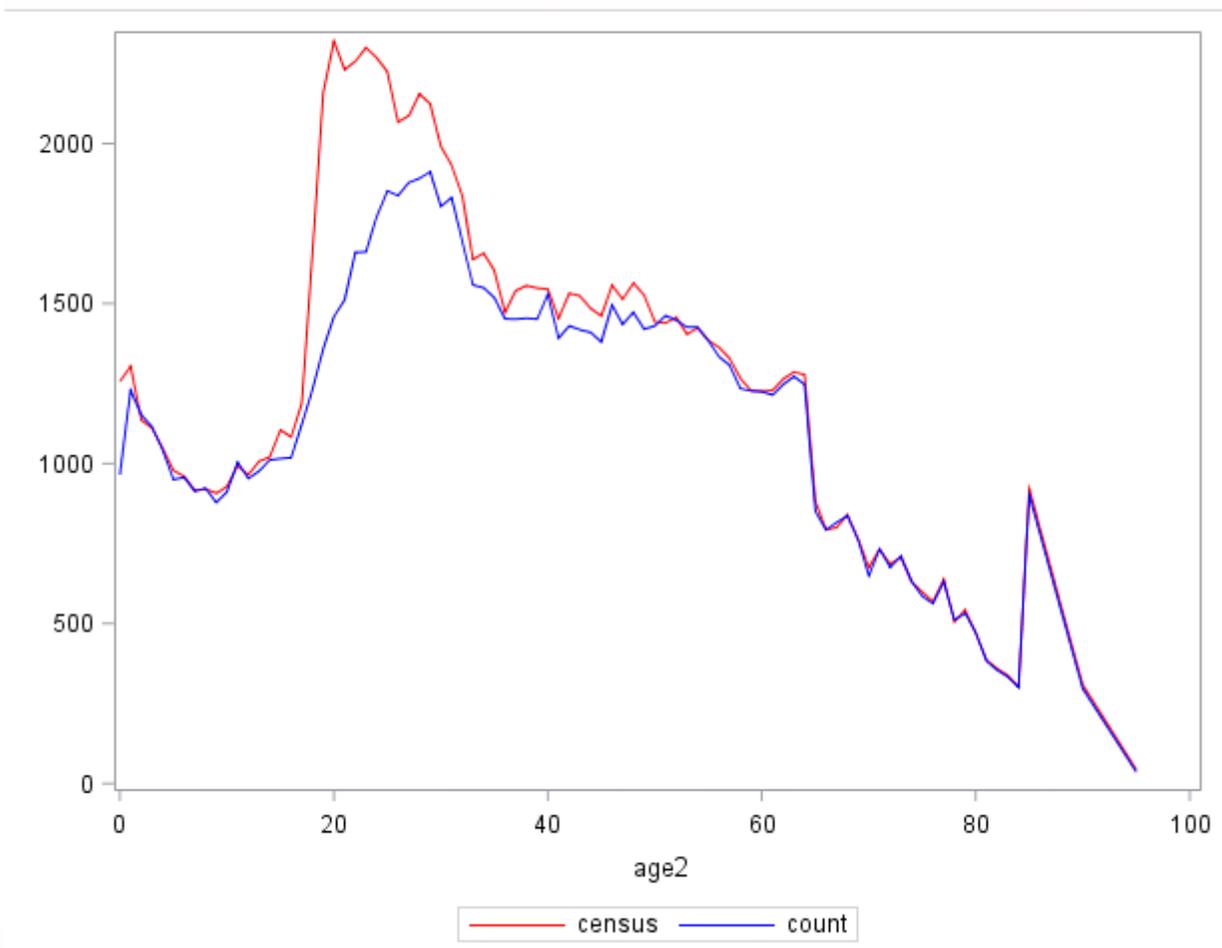
- [1. Introduction](#)
- [2. Operation context and administrative data collection](#)
- [3. Communication with data supply partners](#)
- [4. Quality assurance principles, standards and checks applied by data suppliers](#)
- [5. Producer's quality assurance investigations and documentation](#)

(2) Data Quality

- Meeting with data suppliers –
 - background to the data
 - what checks do they do?
- Our own checks:
 - Missing postcodes / do postcodes map to a council area?
 - Check age – sex distribution
 - Check length of first names and last names
 - Compare with previous years

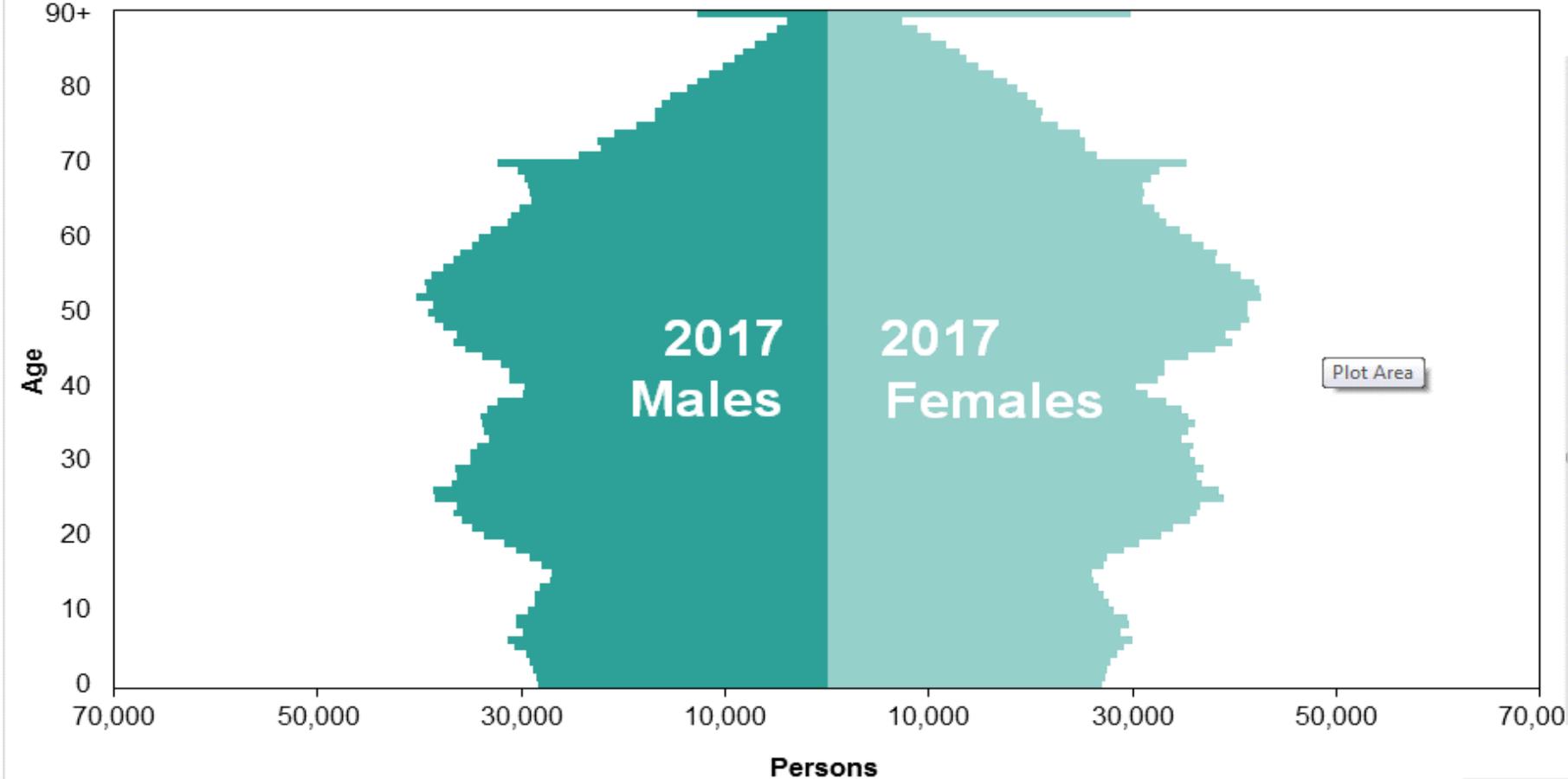
(2) Single Year of Age – Compared with 2011 Census

LA2=Aberdeen City



Population and Household Estimates

Population pyramids of Scotland, 1981-2039



Population Estimates



1. **How do we combine datasets to ensure under-coverage** (if possible).
2. Begin with population counts using 2011 data.
3. Use dual system estimation to adjust for under-coverage and estimate the complete population.
4. Compare with the 2011 Census.
5. Once working, “productionise” for datasets from 2016 through to 2021.
6. Compare with Census 2021 estimates.

Household Estimates



1. Trickier than population estimates. Most datasets relate to individuals and not their relationships with others.
2. Relies on mapping address to the UPRN (unique property reference number)
3. Address information on Electoral Register, Health Activity and Vital Event datasets.
4. These will be occupied property estimates, rather than household estimates.
5. Follows from the population estimates work.

Data Processing Tasks

- What tasks are worth taking forward into the 2019 Census Rehearsal?
- Testing the following work packages to look at the size of the problem, and time required to remedy:
 1. Whole Census De-duplication
 2. Creation of a Synthetic CCS for the 2019 Rehearsal
 3. Matching 2011 Census and 2011 CCS within 20 working days.
 4. Resolving multiple responses within a household.

Data Processing Tasks

- Testing the following work packages to look at the size of the problem, and time required to remedy (continued):
 5. Item level quality assurance using dates of birth from admin data
 6. Test of using admin data to help with placement of skeleton records, created from estimation and adjustment process
 7. Test of Automatic Linkage methods for large datasets.
 8. Removing false persons – answers is the a real person when limited info has been provided.
- Beneficial tasks may be taken forward to the October 2019 rehearsal.

Summary

- Gained access to a wide variety of datasets covering a significant proportion of the population.
- Commencing quality assessments of these datasets
- Testing work packages for how admin data can help with the Scotland's Census 2021



Contact Details

- Dr. Andrew Waugh
 - Andrew.Waugh@nrscotland.gov.uk