Scotland's Census

From paper and internet to a final number (and then detailed outputs)

Head of Downstream Processing Unit, October 2012

National Records of Scotland

Scotland's Census 2011
Shaping our future

Overview

- Census taken on 27 March 2011

- Roughly 80% paper returns, 20% internet.

- To arrive at a population figure we:
  - Capture and clean the data
  - Impute missing characteristics
  - Estimate the returns we didn't get
  - Derive variables for output
  - Assign output areas
  - Disclosure Control of the data

National Records of Scotland

Scotland's Census 2011
Shaping our future

Development of methods

▶ Developed in close consultation with Office for National Statistics (ONS), Welsh Assembly Government (WAG) and Northern Ireland Statistics and Research Agency (NISRA)

▶ Allows harmonised outputs

▶ Implementation by National Records of Scotland (NRS), but making use of ONS algorithms and code where possible.

National Records of Scotland

Scotland's Census 2011
Shaping our future

Capture and Coding

▶ Scanning / Operators

- All tick boxes and text fields captured as text
- Questionnaires guillotined and scanned
- Hundreds of operators
- Questionable fields flagged to operators
- Quality assurance samples drawn and checked

Data Cleaning – Initial Validation

▶ Load and Validation – right types of values/ranges etc
  – Check data received as expected
  – Load into Small Area Statistics (SAS) database
  – Referential integrity
  – Range checks

▶ Remove false Persons – (2 of 6 rule)
  – Occur due to: crossings out/mistakes or dust on scanner
  – Reject person records without a response to at least 2 of:
    • name
    • sex
    • marital/civil partnership status
    • date of birth

National Records of Scotland

Scotland's Census 2011
Shaping our future

Data Cleaning – Multiple Responses

▶ Can occur due to:

  – Internet and paper returns from same household

  – Two paper returns from same household

  – person filling in details twice

  – person on both household and individual forms

▶ Identify which case then

  – decide which is 'best' response (rules)

  – merge data where appropriate

National Records of Scotland

Scotland's Census 2011
Shaping our future

Data Cleaning – Filter rules

▶ Not everyone should answer every question, e.g. own accommodation (skip landlord question), born in UK (skip date of arrival) under 16 (skip employment questions)

▶ Resolve inconsistent responses

▶ Deterministic?

▶ Which response do we believe?

National Records of Scotland

Scotland's Census 2011
Shaping our future

Imputation (1)

▶ Some records have missing/inconsistent data

▶ Probabilistic approach

▶ Requires complex relationships between members of the household to be analysed

▶ Missing and inconsistent responses

Imputation (2)

▶ Canadian Census Edit and Imputation Software (CANCEIS)

▶ Donor imputation

▶ Minimum change

▶ Decision Logic Tables (DLT)

▶ Deterministic edits?

Coverage matching and estimation

▶ Missing households and people

▶ Census Coverage Survey (CCS)

▶ Match Census and CCS records - automatic and clerical

▶ Dual systems estimation

▶ Regression estimator

▶ Age-sex groups by local authority

▶ Overcount?

▶ Estimates Quality Assured against admin sources

National Records of Scotland

Scotland's Census 2011
Shaping our future

Coverage adjustment

▶ Produce consistent individual level database

▶ Add missed households and individuals

▶ Use known gaps where possible

▶ Maintain consistency with surrounding area

▶ 'Skeleton records'

National Records of Scotland

Scotland's Census 2011
Shaping our future

First release

▶ 5 year age bands, by local authority, by gender

Post-Coverage Imputation

▶ We need to fill out realistic characteristics for the skeleton records

▶ Use CANCEIS

Derive complex variables

▶ Remaining variables for outputs, e.g.

- household composition algorithm

- dwellings

- occupation

- industry

Output area creation

► Lowest geographical level of unrestricted data release

► Working on a principle of minimum change from 2001

► Working closely with NRS Geography

Disclosure control

▶ Protect individual-level data by introducing uncertainty

▶ Assuming pre-tabular either over-imputation or record swapping

▶ Level to be decided (and not made public)

▶ Balance between protection and utility

National Records of Scotland

Scotland's Census 2011
Shaping our future

Publication and Dissemination

▶ Phased releases

▶ Increasing detail

▶ Thematic outputs etc

Thank you