



SURVEY FUTURES

SURVEY DATA COLLECTION
METHODS COLLABORATION

Enhanced sampling frames for general population surveys

Paul Smith, University of Southampton

Anna Keyes, ScotCen



Introduction

- Subproject 1 of Research Strand 1
- Feasibility of using information from administrative sources to enhance sampling frames
 - facilitate online approaches for surveys
 - more cost-effective data collection (reduced screening)
 - improved inclusivity

Literature review

- Specific to UK general population sampling frames
- What frame is most widely used in surveys? (Postcode Address File)
- Are there other easily accessible options? (AddressBase Premium)
- How can these frames be enhanced? (Admin data)
- Could administrative data be used as a general population sampling frame?
- What is the basis for evaluating sample frame quality?
- Where are the evidence gaps and future opportunities?

Practitioner guide

- Advice to survey researchers and statisticians on selecting a general population sampling frame, based on most up-to-date evidence:
- Which options are readily available?
- What are their known costs and benefits?
- Is it worth enhancing a sampling frame with administrative data?
- How straightforward would that be?
- Could administrative data be used as a sampling frame?
- Case studies

Additional data

- Long history of adding *aggregate* data to frames
 - aggregate census information linked to PAF
 - Wealth & Assets Survey oversampled areas with more self-assessment tax
- Potential to add microdata
 - linkage to addresses
 - consistency of people at an address

Opportunities

- Reducing costs
 - reduced deadwood cases
 - improved design for subpopulations
 - classifiers for stratification
 - variables as predictors – casewise accuracy helpful but not necessary
 - improved calibration
 - nonresponse compensation
 - accuracy
- Names (unlikely to be available under any current access provisions)
 - named contacts – up-to-dateness?
 - within-household sampling

Challenges

- Access
 - legal constraints on access
 - mechanisms for specific datasets – time-consuming
 - access for sampling not demonstrated in general
 - data owner approval
- Ethics
 - acceptability testing, public interest test
 - ethics vs sampling
- Linkage
 - address-based or person-based?

Sampling facility – rationale

- Access to administrative data challenging
- Linkage of multiple administrative datasets challenging
- Security of linked data critical

- Government already progressed down this route... *but*
 - ONS currently focused on core activities
 - upfront investment in infrastructure

Discussion points

- Evaluation of PAF, AddressBase and RDMF (or similar) should be built into Census 2031 activity
 - golden opportunity for frame evaluation
- Making the case for a “sampling facility”
- Does the enhanced frame deliver enough benefit
 - cost/benefit
 - coordination
 - better information for subpopulations