SURVEY FUTURES

SURVEY DATA COLLECTION METHODS COLLABORATION

Best Practice Workshop: Enhanced sampling frames for general population surveys

University of Essex

University of Southampton

UKRI Economic and Social Research Council

NCRM National Centre for Research Methods

Office for National Statistics

UCL

Warwick The University of Warwick

Manchester 1824 The University of Manchester

City University of London Est 1894

National Centre for Social Research

LSE The London School of Economics and Political Science

Ulster University

Unil UNIL | Université de Lausanne

Ipsos

Kantar Public

Outline

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Introduction
- UK General Population Sampling Frames: strengths, weaknesses, & future opportunities – Anna Keyes (ScotCen)
- Opportunities and challenges of sampling with administrative data – Paul Smith (University of Southampton)
- What the RDMF could do for surveys in the UK: ambition and reality – Gerry Nicolaas (NatCen)
- Discussion

University of Essex
University of Southampton
UKRI Economic and Social Research Council

Introduction

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- 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

- Survey Futures phase 2 project on Reference Data Management Framework

University of Essex
University of Southampton
UKRI Economic and Social Research Council

National Centre for Social Research

UK General Population Sampling Frames: strengths, weaknesses, & future opportunities

11th June 2026

Dr. Anna Keyes, National Centre for Social Research

Background

- SP1 includes a literature review, survey practitioner guide, and paper
- I’m mainly covering the practitioner guide here

Scope of 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?

Scope of 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

Sampling frame quality considerations

- Whether it uses multiple sources of information
- How frequently it is updated
- Estimated under-coverage rate
- Which sub-groups are disproportionately under-represented
- Estimated over-coverage (ineligible/duplicate) rate
- Accuracy of content, sufficient to enable contact

Measuring sampling frame quality

- Deadwood rates for face-to-face surveys
- England & Wales Census 2021 quality assurance of Addressbase Premium
- Comparison with mid-year population estimates, Census data

Readily available sampling frame option 1: Postcode Address File

- The most popular UK general population sampling frame for 40 years and counting.
- Royal Mail: ‘the UK’s most up-to-date and accurate address database’.
- Regularly updated, affordable to access via licence, clean & trustworthy address data.
- Addresses only, cannot systematically exclude non-residential.
- Considerable evidence on dealing with these limitations via weighting.

Readily available sampling frame option 2: AddressBase Premium

- For GB, PAF plus additional data: National Land and Property Gazetteers and Ordnance Survey Map.
- Separate NI version, Addressbase Islands.
- Addresses and unique property identifiers, with classification into residential and commercial uses.
- Advantages of PAF, plus means to drop non-residential addresses.
- More opaque (higher) cost than PAF, limited validation of classification.

AddressBase Premium as England & Wales 2021 Census frame

- AddressBase initially developed for 2011 Census.
- Quality assurance for 2021 E & W Census combined Addressbase with council tax records. Of total addresses:

  95% classified as residential by Addressbase had council tax records

  4% classified as residential by Addressbase but no council tax record

  0.6% not classified as residential by Addressbase but had council tax records

Adding administrative data to PAF or AddressBase

Literature is focused on:

- Linking admin data to survey responses for enhancement or validation,
- Linking admin data to address frames for population estimation,
- Linking admin data to sampling frames for subgroup identification.

Health records are the most commonly used form of UK administrative data for these purposes.

Adding administrative data to PAF or AddressBase

Lessons from literature and case studies:

- Accessing administrative data can be lengthy process, varying across public bodies and nations of the UK.
- Data matching on addresses is difficult, protracted, inexact due to lack of unique identifiers shared across data sources.
- Methods are advancing, successful examples of screening PAF for particular age groups using health records, e.g. Scottish Health Survey child boost.

Using administrative data as a sampling frame

- Unusual to use admin data alone as a *general population* sampling frame, much more common as a specific subgroup sampling frame.
- Lothian Health Survey 2023 used Community Health Index records as general population sampling frame.
- Enabled stratification by age (16 to 24 and over-25s), named sample.
- Data quality issues, comparison with Scottish Census showed concerning divergence in subgroup population estimates.

Using administrative data as a sampling frame

Lessons from case study:

- Sample design must incorporate outcomes of a sampling frame quality evaluation.
- Direct data access isn't essential; data controller could draw sample from specification then share code for quality assurance; anonymised sampling frame with unique IDs another option.
- Processes needed to clean administrative data, removing duplicate or ineligible cases and ensuring addresses are accurate (can be contacted).

Key evidence gaps

- Is the additional cost of Addressbase worth the saving of lower ineligibility rates? More validation needed.
- Could adding administrative data to PAF or Addressbase offer significant improvements in sample stratification on a general population survey? More quality evaluation needed.
- Are administrative databases accurate and comprehensive enough to be used as general population sampling frames? More quality evaluation needed.
- New data sources have future potential but are untested as sampling frames e.g RAPID (DWP and HMRC records). More evidence needed.

How to choose the best general population sampling frame?

Some considerations:

- To date, insufficient evidence to support a wholesale shift away from PAF.
- Reducing ineligibility rates could be higher priority on web or telephone surveys than face-to-face, as harder to control for with weighting.
- Solid evidence for use of admin data with PAF for subgroup boosts.
- Very limited use of admin data as a general population sampling frame, currently risky.

National Centre for Social Research

The Enhanced Sampling Frames practitioner guide will soon be published here.
https://surveyfutures.net/practice-guides/

The Enhanced Sampling Frames literature review is available here.
https://surveyfutures.net/wp-content/uploads/2026/05/working-paper-17-enhanced-sampling-frames-a-literature-review.pdf

National Centre for Social Research

Thank you

T. 020 7250 1866
F. 020 7250 1524
E. info@natcen.ac.uk
W. www.natcen.ac.uk

Registered Office

35 Northampton Square
London EC1V 0AX

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Opportunities and challenges of sampling with administrative data

Paul A. Smith, University of Southampton

University of Essex
University of Southampton
UKRI Economic and Social Research Council

Outline

- History
- Opportunities of administrative data
- Challenges of administrative data
- “Sampling facility”
- Discussion points

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

University of Essex
University of Southampton
UKRI Economic and Social Research Council

History

- 50s-80s: Electoral register
  - built-in lags
  - names
- Postcode Address File
  - introduced in 1983 for LFS
  - extensive assessment
    - several evaluations based on comparisons with census data (“nonresponse link study”)
    - split sample test on British Social Attitudes Survey
  - stable? No evaluations since 1990s

Additional data

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- 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
- Predictors vs ‘truth’

Opportunities – costs

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- 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

Opportunities – names

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- names are personal information – not currently available
- named contacts
  - if names are *current* could improve response
  - if names are *outdated*, likely to depress response
  - no current empirical evidence (?)
  - impact depends on balance of names currency
- within-household sampling
  - can select named person
  - accuracy of names affects control of within-household samples

Opportunities – characteristics from names

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Names carry information on
  - sex
  - ethnicity
  - age
- Machine learning models improving prediction
  - still biases from census data -> no replacement
- Creates additional variables for stratification. Predictors, not true values

Opportunities - characteristics

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Observed characteristics used in stratification
- Linking to addresses or people
- Early Life Cohort used birth records linked to maternity records
  - ethnicity

Challenges – access

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Legal constraints on access
  - SRSA 2007 and Digital Economy Act 2017 *enabled* data sharing
  - data available for research in the public interest in secure facilities
  - access for sampling not demonstrated in general
- Data owners
  - case for use – acceptability testing, public interest test
- Mechanisms for specific datasets
  - time-consuming

Challenges – ethics

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- (Mostly) no opt-out for inclusion in administrative data
  - no informed consent over inclusion
  - public acceptability of use of admin data in sampling
- Sampling
  - will inconvenience sampled people (households)
  - design aimed at differential inclusion -> no strict equality
  - balance the inconvenience to sampled people (households) against the benefit of better (lower variance) and/or more efficient (lower cost) survey outputs
- Constraints
  - minimum subpopulation size

Challenges – linkage

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Different levels of linkage
  - address-based – addition to address frame
  - person-based – all information linked to same person, and person linked to address
- Permissions
  - ideally informed consent, but not practical over admin datasets
  - public acceptability

Sampling facility – rationale

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

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

- Government already progressed down this route
  - administrative data linkage to give demographic index
  - research on administrative data census
  - comparison with census outputs

Sampling facility – realism

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- The most practical place to gather data together in an enhanced frame is within government
  - ONS best placed because of its existing research and data holdings
  - would require substantial development

- ONS currently focused on core activities
- upfront investment in infrastructure

- Following slides are *my* assessment
  - additional development needed

Sampling facility – legal

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- SRSA 2007

22 Statistical services

(1) The Board may provide statistical services to any person in any place within or outside the United Kingdom.

(2) The services which may be provided under this section include in particular—

(a) providing information, advice and technical assistance in relation to statistics;

23 Statistical research

The Board may promote and assist statistical research, in particular by providing access (where it may lawfully do so) to data held by it.

- approval from data owners explicitly for 22

Sampling facility – ethics

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Clearly not appropriate to release characteristics information
  - including names
- Most that could be released would be sample of addresses
  - obtained by a probability sampling procedure
    - good practice
    - fairness
    - disclosure protection
  - no sampling of very specific populations (ie restricted to general population sampling)
    - minimum population size
    - possibility of sample coordination

Sampling facility – population data

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- As well as selected sample, require supporting information
  - population totals from frame
  - calibration totals for relevant subpopulations, by frame variable
- Additional consistency between surveys

Sampling facility – costs

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Set-up costs and running costs
- Set-up: Initial work required on
  - data access, permissions
  - ethics and review
  - security
- Infrastructure: Hopefully some existing work on admin data provides a basis for frame management, linkage, quality evaluation (RDMF)
- Ongoing costs for frame maintenance and drawing of samples recovered from fees for sample provision
  - cost-benefit
    - better accuracy/lower cost
    - no need for in-house sampling facilites?

Discussion points

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- 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

SURVEY FUTURES

SURVEY DATA COLLECTION METHODS COLLABORATION

What the RDMF could do for surveys: Ambition and Reality

Gerry Nicolaas, Zac Perrera & Nick Waugh, National Centre for Social Research

In collaboration with the Office for National Statistics

Background:
Growing pressure on UK surveys

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Response rates down
- Costs up
- Demand for more timely, inclusive and granular data

University of Essex
University of Southampton
UKRI Economic and Social Research Council

The problem

- No UK population register
- UK relies on address-based sampling for general population surveys
- Lack of person-level information
- Limited use of administrative data in design

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

University of Essex
University of Southampton
UKRI Economic and Social Research Council

What is the Reference Data Management Framework (RDMF)

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- A centralised infrastructure developed by the ONS
- Designed to support secure and consistent linkage of admin data
- Four linked indexes:
  - Demographic
  - Business
  - Location
  - Classification
- Architecturally distinct from a population register
  - Unsuitable for sampling of individuals given design, coverage & governance constraints

University of Essex
University of Southampton
UKRI Economic and Social Research Council

Research questions

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- RQ1 - How could the RDMF be used to **improve** the end-to-end survey data collection process?

- RQ2 - What are the **constraints and possible solutions** for using the RDMF in the design and implementation of social surveys?

Analytical approach

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

TOTAL SURVEY QUALITY	
TOTAL SURVEY ERROR	**NON-STATISTICAL QUALITY**
Representation • Target population definition • Frame construction • Sample selection • Fieldwork • Weighting	• Relevance • Timeliness • Credibility • Coherence/comparability • Usability/accessibility
Measurement • Operationalisation • Questionnaire design • Mode selection • Pre-testing • Processing	

The Ambition

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

End-to-end survey process	Non-statistical quality
Better address coverage than PAF	More relevant data by improving inclusivity
Clearer identification of communal addresses	Enhanced credibility through more representative data
Individual-level data could strengthen sample design	Better coherence across sources
More cost-effective fieldwork (tailored/adaptive designs)	Better comparability across datasets
Improved data processing	
Improved weighting (richer auxiliary data)	
Reduced questionnaire length	
Less reliance on error-prone self-reports	

The Reality

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Demographic Index
  - Undercoverage of those with weak administrative footprints
  - Overcoverage of those who have died or emigrated
  - Update lags
- Linkage quality still being assessed
- Conceptual mismatch between admin definitions & survey concepts
- Governance and access limits

Most practical opportunities

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Use the Location Index
  - Better address coverage than PAF
  - Fewer ineligible cases
  - Better visibility of communal settings
- Link auxiliary data to sampled addresses via the RDMF
  - To produce more inclusive samples
  - To inform more effective non-response strategies
  - To reduce respondent burden by replacing survey questions with equivalent variables from linked administrative records
  - To improve quality assurance, data processing, and weighting

Conclusion

- Promising, but not plug-and-play
- Best short-term use: Location Index
- Bigger gains possible, but conditional
- Needs evidence, governance, investment

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

University of Essex
University of Southampton
UKRI Economic and Social Research Council

Rethinking the Role of Admin Data

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

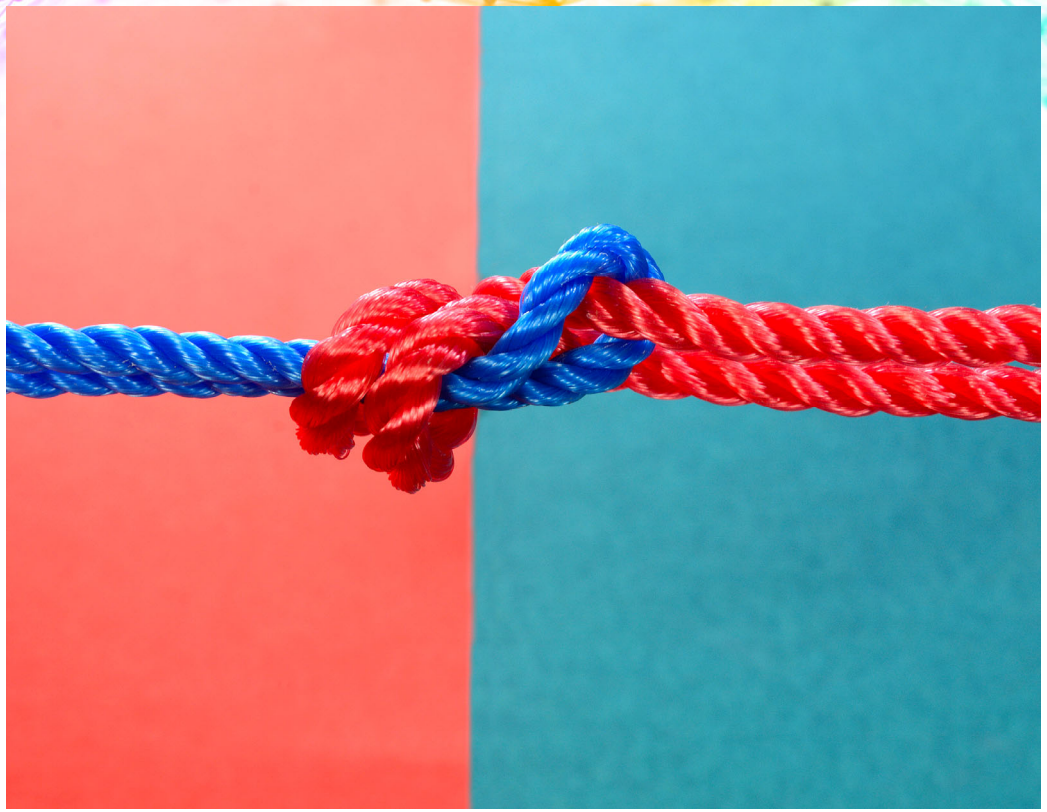
- Admin data and survey data do different things
- Surveys remain essential — admin data cannot replace them
- Admin data can improve and support surveys

University of Essex
University of Southampton
UKRI Economic and Social Research Council

The fundamental question

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

How do we design an evidence system in which surveys and administrative data are mutually reinforcing - without sacrificing security, trust and legitimacy?



University of Essex
University of Southampton
UKRI Economic and Social Research Council

Thank you

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Publications at https://surveyfutures.net/publications/

- ONS (2026) **Reference Data Management Framework project - what is the RDMF?**
- Perrera, Nicolaas & Waugh (forthcoming) **How could the Reference Data Management Framework improve survey quality?**
- Cornick (forthcoming) **Using the ONS Reference Data Management Framework: a cost-quality comparison**
- ONS (forthcoming) **Q&A on how the RDMF could potentially be used to improve the design and cost-effectiveness of the Transformed Labour Force Survey**

University of Essex
University of Southampton
UKRI Economic and Social Research Council

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Questions and discussion

University of Essex
University of Southampton
UKRI Economic and Social Research Council