SURVEY FUTURES

SURVEY DATA COLLECTION METHODS COLLABORATION

Combining probability and non-probability samples

David Hussey(NatCen), Matthew Wong (NatCen), Dhru Shah (NatCen)

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

Introduction

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Sub-project 2 of Research Strand 1
- Methods for combining - or integrating - probability and non-probability samples.

- In practice this means:
  - Combining parallel probability and non-probability surveys
  - Adjusting non-probability survey data using a “reference” probability survey

Background

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Probability samples
  - Still the gold standard
  - Increasing challenges - response rates, costs, demand changes

- Non-probability samples
  - Advantages: speed, cost, scale
  - Problems: selection bias, coverage error

- Integration offers potential to combine strengths of both
  - Proviso: if able to deal with selection bias

What did we do?

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- **Evidence review**
  - Evidence from the UK and other countries
  - Focus on empirical studies – those that combine / integrate samples

- **Practitioner guide**
  - Advice to survey statisticians, analysts
  - Covers methods available, when to use them and how (in steps)

- **Working paper**
  - Used existing data collected in two parallel surveys (originally combined using PSW)
  - Aim: reduce selection bias further using additional covariates

Motivations

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- **Primary goal**: reduce selection bias in a non-probability sample

- Additional Motivations
  - Increase overall sample size to improve precision
  - Overcome issues with coverage
  - Improve estimation from small probability samples

Methods

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- **Design-based** - produce weights for a non-probability sample to align it with a “reference” probability sample.
  - Key Methods: Propensity Score Weighting, Calibration

- **Model-based methods** - use statistical models to predict survey outcomes in nonprobability sample -> apply to probability sample.
  - Key Methods: Bayesian Inference, Mass Imputation

- **Hybrid approaches** combine elements of both approaches.
  - Key Methods: Composite estimation, Doubly robust

Strengths and limitations

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- **Design-based**
  - Pros: intuitive, accessible, single set of global weights
  - Cons: ineffective with small probability samples, struggle with high "dimensionality"

- **Model-based**
  - Pros: more flexible, can use small probability samples & incorporate "high-dimensional" covariates
  - Cons: less accessible, computationally intensive, sensitive to model misspecification

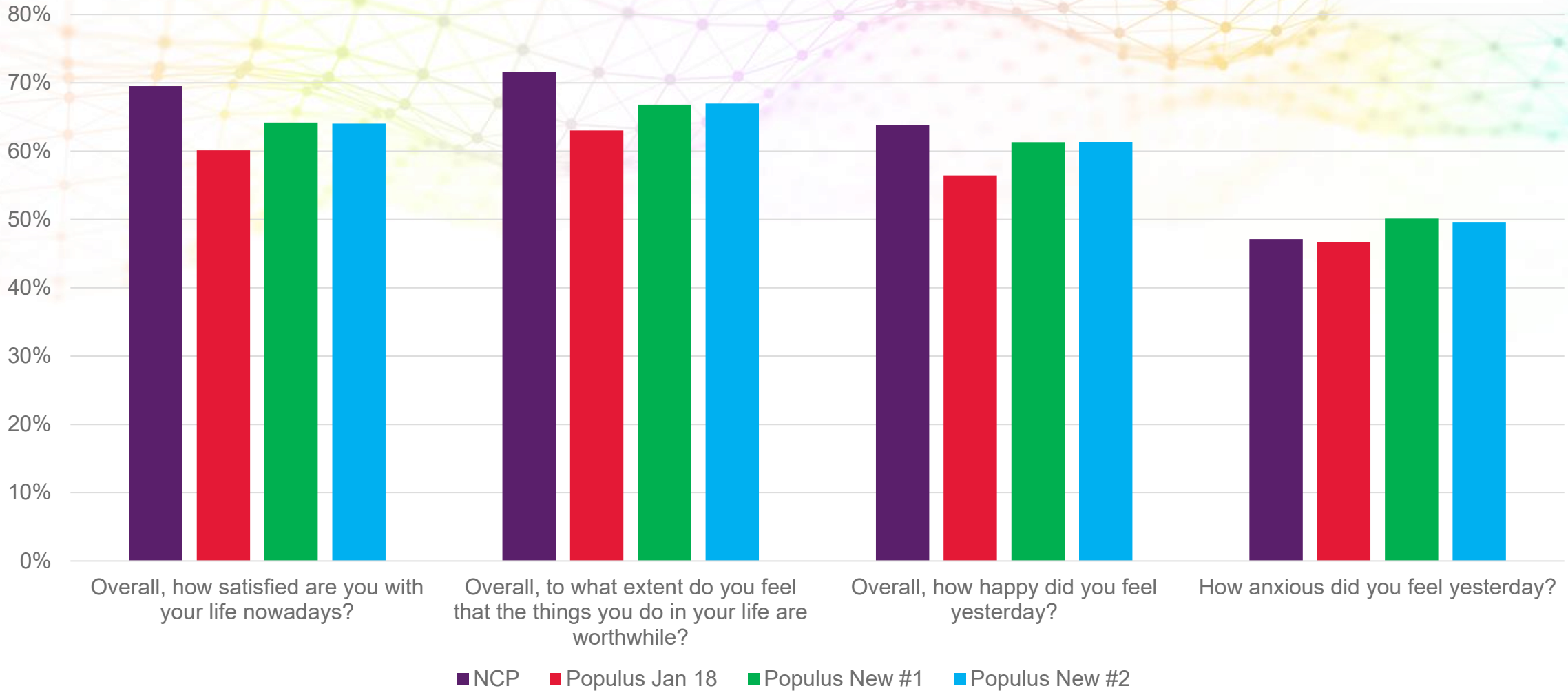
Key learnings

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Wide range of methods available. Recent growth of Bayesian / machine learning.
- Key conclusion: relative paucity of evidence on performance in real world settings.
- Different methods suit different data structures.
- Method secondary compared to the availability of useful covariates.
- Newer methods tend to out-perform traditional.
- Future research: plenty of scope.

Case study

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION



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

Discussion points

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

- Why are these methods not used more widely, especially in the UK?
- What are the barriers to use?
  - Knowledge gap
  - Statistical skills required
- Do they have potential to expand the use of non-probability surveys in the UK?