MIXED MODE RESEARCH TO INFORM THE LTSS

Rachel Ormston, Chris Martin, Laoise Rogers (Ipsos Scotland); Tom Huskinson, Eileen Irvin, Emma Rimmington (Ipsos Survey Research Methods Centre); and Peter Lynn (University of Essex)

6 December 2024

Presentation outline

Context and aims

Methods and data

What do we mean by mixed mode?

Key messages from the research

Questions and discussion

CONTEXT AND AIMS

The Scottish Government's Long-term Survey Strategy (LTSS)

- Sets out vision and plans for SG population surveys, focusing on:
  - Scottish Health Survey (SHeS)
  - Scottish Crime and Justice Survey (SCJS)
  - Scottish Household Survey and House Conditions Survey (SHS/SHCS)
- Four LTSSs to date – latest 2018-22
- Emphasised analytical potential, survey quality and value for money, but noted risk from falling response rates
- Next LTSS delayed by pandemic but in development – research commissioned to inform it.

The Scottish Government's 'flagship' gen pop surveys

SHS

- Est 1998
- 10,500 ints / year
- Wide-ranging topics
- Household interview + 'random adult'
- Travel diary
- Physical survey of property

SCJS

- Est 1994
- 5,500 ints / year
- Key info on victimisation + views on CJ system
- Includes victim form + self-completion for sensitive crimes

SHeS

- Est 1995
- 5,000 adult + 2,000 children / year
- Health and health-related factors
- Biological measures (sub-sample)

The wider context

- Pressure on public sector budgets
- Concern about declining response rates across F2F surveys
- Increasing internet access across population (with important exceptions)
- Major disruption of Covid-19 to face-to-face surveys
  - All F2F fieldwork halted – surveys forced to experiment with other modes
  - Challenges maintaining F2F interviewer panels

Acceleration of pre-existing trends (budgets, RRs) => increased impetus to explore alternatives

The Scottish Government commissioned this research in order to explore the **potential benefits and risks of mixed mode approaches** (i.e. collecting data via another mode, such as online or telephone, or by a combination of modes) in the specific context of the three main Scottish Government general population surveys.

Key points to keep in mind

1

Focuses on the three main **cross-sectional, random probability** general population surveys (SHS, SHeS, SCJS)

2

Intended to inform the LTSS and associated discussions – but **not** to make recommendations

3

Scope is still huge – 200-page main report, so this presentation is necessarily selective.

METHODS AND DATA

Methods and data

Interviews with 25 stakeholders from within and outwith the Scottish Government

Desk-based review of key literature on mixed mode surveys

Interviews with 23 survey experts from outside Scotland

Review of 21 key surveys from the UK and further afield

Stakeholder workshop

Structure of summary report

- Key messages
- Introduction, context, scope, methods, definitions
- Key issues, mitigations & trade offs:
  - Representation: coverage and sampling
  - Representation: nonresponse
  - Measurement error & mode effects: impacts on different question types
  - Implications of mode for data collection options
  - Impacts of changing or mixing modes on time series
  - Survey quality metrics
  - Financial & resource considerations
- Administrative data
- Good practice when considering mode change

WHAT DO WE MEAN BY 'MIXED MODE'?

‘Mixed mode’ is used to describe surveys that combine two or more ways of collecting data from respondents, such that different respondents complete either the survey as a whole (or the same section of a survey) by two or more different modes.

Mixed mode surveys can vary in terms of …

1

Mode of invitation (not necessarily the same as mode of completion)

2

Mode of participation or completion (11 possible combinations of f2f, telephone, paper, and web)

3

Sequencing (concurrent or sequential)

Different designs in practice

Crime survey for England and Wales

F2F 1st wave, with longitudinal telephone follow-up. Exploring feasibility of web.

Health Survey for England

F2F (2022), but considering web or web + telephone designs (both with F2F nurse visit)

English Housing Survey

F2F but testing push-to-web with F2F follow-up

National Survey for Wales

Push-to-telephone, with F2F knock to nudge to increase response.

Transformed Labour Force Survey

Push-to-web with telephone matching and targeted F2F knock to nudge.

KEY MESSAGES FROM THIS RESEARCH

One key take-away: There is no simple formula for deciding whether/how to transition to a mixed mode design – there are always trade-offs, and some are more difficult to mitigate.

Key messages re. trade-offs

Need to understand nonresponse bias

Weighting etc. can reduce, but not eliminate this.

Each mode has likely nonresponse biases

F2F

Better: people from deprived areas; lower literacy

Relatively weaker: working households, flats

Web

Underrep: older people; digitally excluded; deprived areas; ESOL; renters; large HHs.

Overrep: engaged

Phone

Similar biases to web compared with F2F

Paper

Underrep: low literacy, ESOL.

Better for: older people, compared with web

Potential mitigations

- Multiple contact strategies, ideally tailoring to groups known to be less likely to respond
- Monetary incentives (non-monetary might increase nonresponse bias)
- Weighting.

Remaining issues and trade-offs

- Practical limitations to tailoring of contact strategies (sample frames = address-based)
- Unconditional incentives more effective – but more 'wasteful'
- Weighting can't eliminate nonresponse bias (especially where not demographic – e.g. more engaged more likely to respond)

Key messages re. trade-offs

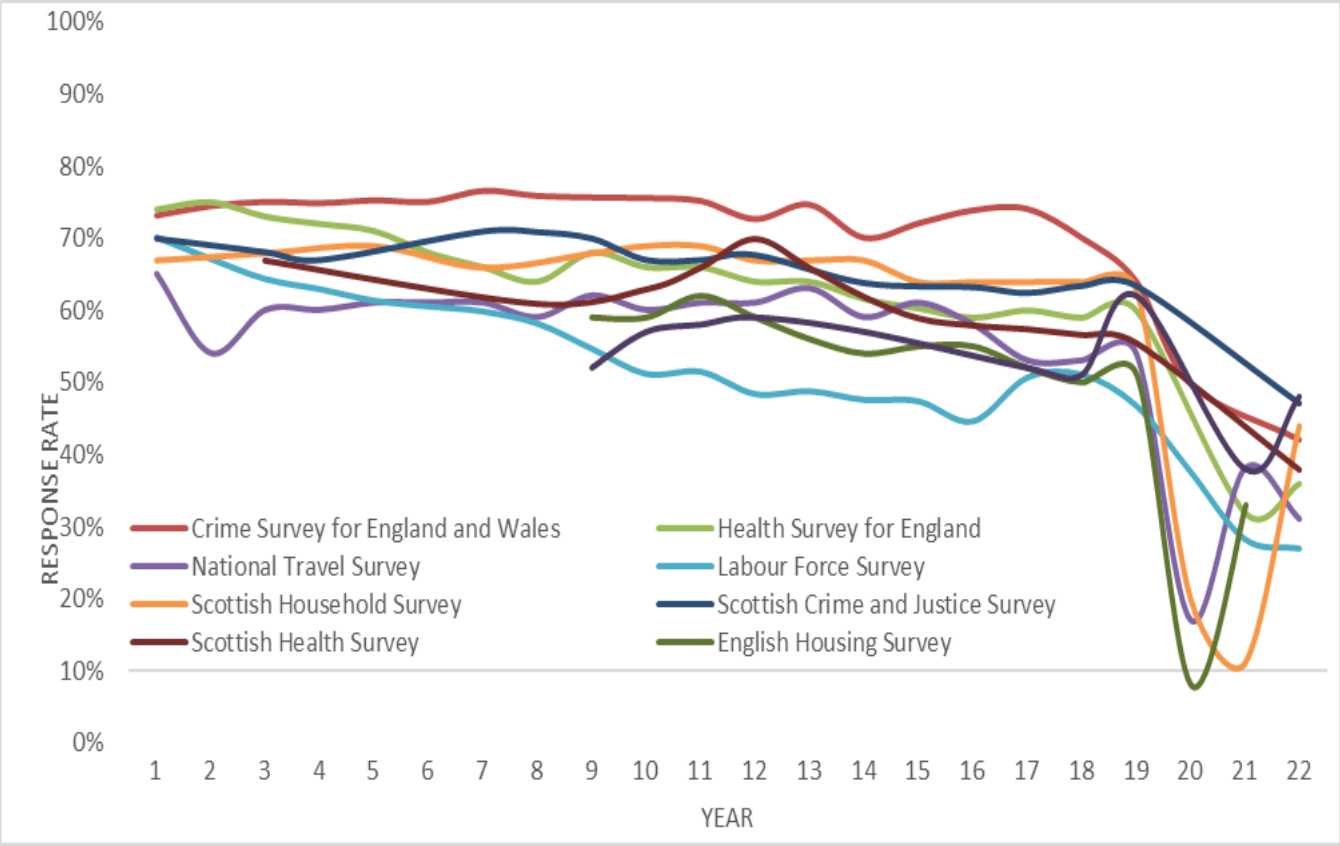
Need to understand nonresponse bias

Weighting etc. can reduce, but not eliminate this.

Need to consider alternative quality measures to response rates

(or determine what is an acceptable RR).

Why do we need to consider alternatives to RR?



- Downward trend - we don't know what the 'new normal' is (for any mode)
- RRs matter insofar as they impact on nonresponse
- BUT association between survey RRs and NRB is fairly weak
- And varies between different measures.

Alternatives to response rates

- Focus on nonresponse bias and item nonresponse patterns instead
  - Comparisons of sample profiles vs external estimates
  - Assessing change in item nonresponse (where not expected to change)
  - Draw on existing frameworks (e.g. ONS LFS Performance and Quality reports)

Remaining issues and trade-offs

- RRs are simple – alternatives are not
- Difficult for non-expert users to digest
- Response rates remain important to some data users
- But shared interest in data quality a start point for conversation.

Key messages re. trade-offs

Need to understand nonresponse bias

Weighting etc. can reduce, but not eliminate this.

Need to consider alternative quality measures to response rates

(or determine what is an acceptable RR).

Reducing mode effects may require significant question redesign

(and may highlights issues with existing Qs).

Potential mitigations

- Question redesign (unimode, best practice, optimode)
- Cognitive testing
- Respondent Centred Design and Device Agnostic design principles

Remaining issues and trade-offs

- Mode effects are question-level – so redesign may be a huge undertaking
- Which mode is closer to 'true' value is generally unknowable
- Testing can highlight issues with long-standing questions – redesign may involve 'slaughtering holy cows'
- Mode effects can be reduced but not eradicated (& apply more to certain question type).

Key messages re. trade-offs

Need to understand nonresponse bias

Weighting etc. can reduce, but not eliminate this.

Need to consider alternative quality measures to response rates

(or determine what is an acceptable RR).

Reducing mode effects may require significant question redesign

(and may highlights issues with existing Qs).

Changing mode may mean accepting a break in the time series

Recalibrating everything is not a feasible solution.

Potential mitigations

- Maximise efforts to mitigate mode effects and maintain sample representativeness
- Parallel runs to understand impact

Remaining issues and trade-offs

- Parallel runs are time & resource intensive
- Helps you understand impact of mode change on mode effects/NRB, but doesn't eradicate it
- Recalibration to maintain time series not viewed as feasible or advisable (at scale / at all)
- So trade-off between maintaining trends and other priorities is a fundamental one.

Key messages re. trade-offs

Need to understand nonresponse bias

Weighting etc. can reduce, but not eliminate this.

Need to consider alternative quality measures to response rates

(or determine what is an acceptable RR).

Reducing mode effects may require significant question redesign

(and may highlights issues with existing Qs).

Changing mode may mean accepting a break in the time series

Recalibrating everything is not a feasible solution.

Some additional tasks are especially difficult if move away from F2F

(e.g. physical survey, height & weight, victim form).

Unpacking the financial implications of mode redesign

Drivers of cost

- Sample size
- Recruitment & incentive strategy
- Length
- Extent of piloting/development
- Complexity of sample management, data processing, weighting
- F2F/Tel: reissues
- F2F: clustering

Unpacking the financial implications of mode redesign

Short-term

Question (re)design, testing, // runs, new systems, stakeholder engagement

Medium-term

Depends on combination & other concurrent changes, BUT savings bigger if move to predominantly self-complete.

Long-term

Interviewer costs are biggest cost and likely to go up more. But other costs not static: postal; call blocking tech; incentives

So how on earth do you start?

1

Engage stakeholders as much as possible, as early as possible, as clearly as possible

2

Be clear on the aims of considering mode change – and maintain focus on these

3

Plan to invest in significant development and testing work

4

Consider how ‘future proof’ different options are likely to be

THANK YOU

NAME:
Rachel Ormston

DETAILS:
Rachel.Ormston@Ipsos.com