Office for National Statistics
Swyddfa Ystadegau Gwladol

ONS Plans for Web Data Collection in LFS

Charles Lound (for Salah Merad)

Sample Design & Estimation, Social Surveys

Strategy and Standards Directorate

Talk outline

- Background
  - Current LFS
  - Aims of Electronic Data Collection project
- Some mixed-mode designs
  - Mode effects
  - Estimation methods
  - Other issues
- A cost model
- Plans for a pilot

Outline of current UK LFS

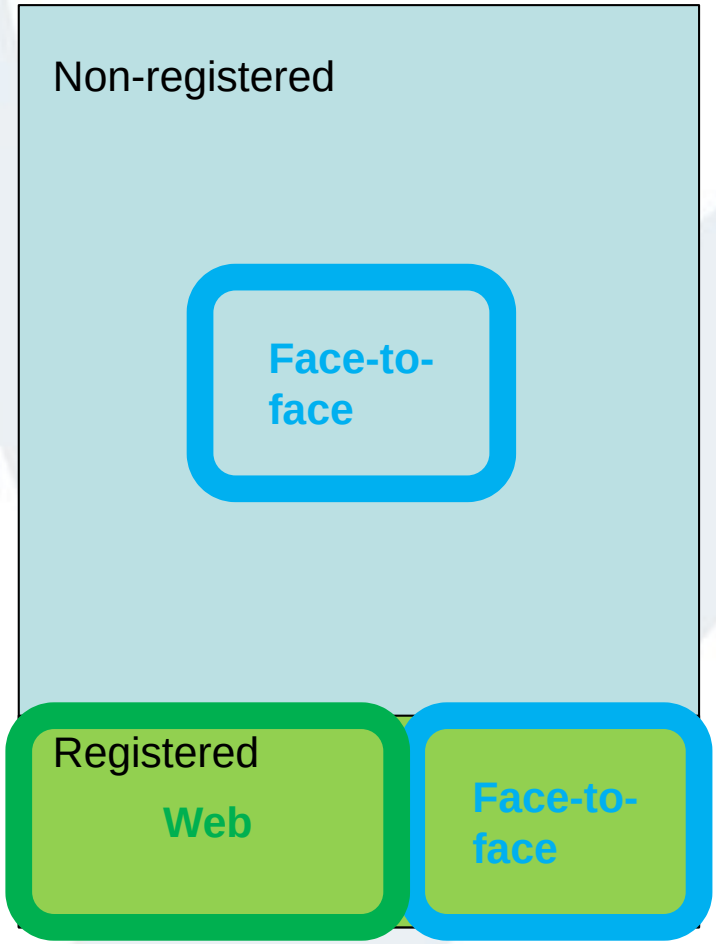
- Single-stage sample of addresses from PAF
  - One household per address
  - All adults in households
- Addresses in sample for 5 waves (consecutive quarters).
  Rotate 1/5th sample each quarter.
- Wave 1 sample of ~10,000 (responding) households per quarter.
- All have interviewer-administered interviews:
  - wave 1: mostly FtF, some Tel
  - waves 2-5: mostly Tel, some FtF
- Estimation: all responses in quarter pooled together. Calibrated to (known) population totals by age, sex and location

Aims of EDC Project

- Want to introduce web option:
  - save money, more efficient, ‘expected’, ‘modern’, less burdensome?, better response?, …
- Introduce (initially) in addition to ‘usual’ LFS:
  - parallel run
  - no damage to ‘usual’ estimates!
  - to assess any mode effect
  - develop ‘best’ estimator
- Later
  - Switch ‘usual’ LFS cases to web, reducing number of FtF/Tel cases.

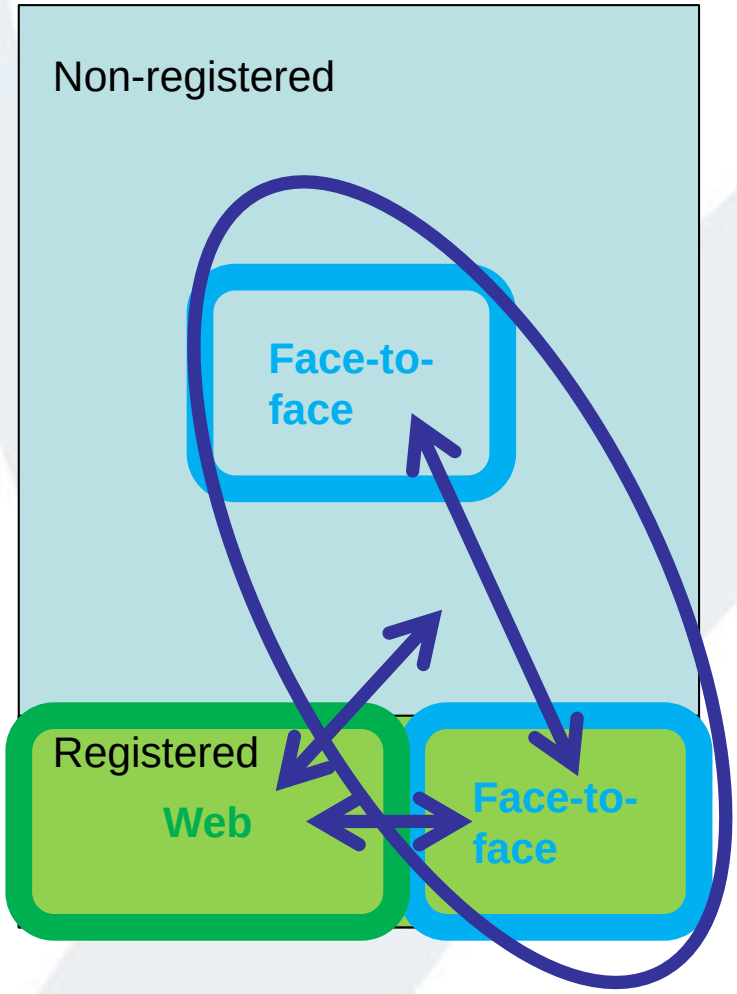
Alternative mixed-mode designs

- Sequential
  - Offer web option
  - Follow-up face-to-face
- Dual-Frame
  - Large sample
  - Ask for web-registration
  - Non-registered
    - Sample for face-to-face
  - Registered
    - Sample some for face-to-face
    - Remainder web



Estimating mode effects

- Compare FtF group with web group
  - Little control over selection effects
  - Large sample sizes
  - Some experience from Opinions Survey pilot
- Compare FtF groups
  - Selection into mode
- Compare responses for registered sample, by mode
  - More control for selection effects
  - But samples sizes may be too small
- Adjustment for mode effects – alternatives
  - Benchmarking to the unbiased estimate (FtF estimate in Dual Design)
  - Unit-level response modification – via regression modelling



Mode effects in *Opinions* Survey (1)

- 2010 online Pilot – November and December
- Demographic questions; some LFS and OPN module questions
- One-person interviewed in each selected HH
- FtF *Opinions* survey as control
- Response for web survey poor in November (8%); better in December (17%) – letter amended
  - 54% response rate to FtF survey

Mode effects in Opinions survey (2)

- Selection effects: more men, more 45 to 64 year old people in web survey
- Logistic regression on employment variable

Covariates	Coefficent for FtF mode	Odds ratio for employment
Mode	-0.22	0.65 [0.54-0.78]
Mode, Age, Sex, Region	-0.22	0.64 [0.51-0.80]
Mode, Age, Sex, Region, HHSize, Marital Status, HRPeducation, Ethnicity, Tenure	-0.12	0.80 [0.62-0.99]

Estimation methods

- Need to adjust for non-response
  - Information would be available for registrants but not for non-registrants
- In Dual Frame design, would need to use a composite estimator as there would be a large variation in design weights
  - The web estimate needs to be adjusted for self-selection and measurement effects

$$\hat{\theta} = \gamma \hat{\theta}_{FtF} + \left(1 - \gamma\right) \hat{\theta}^{*}_{web}$$

Other issues

- What information to collect at registration?
- How to identify non-eligible addresses?
- Multi-households at one address.
- How to handle rotation of sample from quarter-to-quarter.
- Flexibility to allow change of modes during web collection? E.g. re-issues or respondent's choice … not currently planned
- Managing fieldwork
- Cost

A simple cost model

- Cost of letter = £1
- Cost of FtF interview = £25
- Estimated cost of web interview = £1
- Current LFS design – Wave 1
  - Sample size = 16,800
  - Number of responses = 10,000
  - Total cost = £267,000
- Assume three modules in web questionnaires
- Target: achieve 10,000 responses for the modular questions

Cost under a Sequential Design

- Assume 60% FtF response

Web response rate	Eligible set sample	Web responses	FtF responses	Web responses per module	Total responses per module	Cost (£000)
20%	18,300	3,660	8,784	1,220	10,004	242
15%	18,300	2,745	9,333	915	10,248	254
25%	18,300	4,575	8,235	1,525	9,760	229

±6% variation in expected FtF

Cost under Dual Frame design

- Assumed: 80% web response rate for registrants
- Set FtF sampling fraction to 5% to
  - obtain a cost similar to that under Sequential Design
  - obtain a sufficient number of FtF responses

Web response rate	Eligible set Sample	FtF responses	Web responses	Web responses per module	Total responses per module	Cost (£000)
20%	124,000	3,720	18,848	6,283	10,003	236
15%	124,000	3,720	14,136	4,712	8,432	231
25%	124,000	3,720	23,560	7,853	11,573	241

More stable FtF

More volatile

Plans for a pilot (1)

- Refine sampling and develop estimation in 2013-14
- Registration survey pilot planned for 2014
  - Estimate rate of web take up
    - Variation across groups/regions
  - Response to questionnaire after registration
    - Evidence of further selection effects
- Large scale parallel run tentatively planned for 2015
  - Dependent on investment in systems development

Plans for a pilot (2)

- Two possible approaches
  - Current LFS as control group – expensive
  - With a quasi-control group
    - FtF group in Dual design to be of similar size to current LFS size
    - More difficult to estimate discontinuity
- Dual design may be more appropriate for the pilot to estimate mode effects
- Dual design may not be practical to roll out if mode effects are found to be important - the FtF sample would be too small to allow for adjustments in the estimation

Contacts

- Sampling and estimation:
  salah.merad@ons.gov.uk
- Data collection methodology:
  jayne.olney@ons.gov.uk