



Statistics Sweden

Statistiska centralbyrån



The Leverhulme Trust

Stopping rules in a longitudinal survey – impact on cost and survey quality

AAPOR 15th May 2016

Anton Johansson, Statistics Sweden

Peter Lundquist, Statistics Sweden

Gabriele Durrant, University of Southampton



Introduction

- Lower data collection costs
 - Reduce unsuccessful contact attempts
- Constraint:
 - Do not increase bias
- If we can find strategies that also decrease bias, that is a bonus
- In collaboration with Gabi Durrant, University of Southampton

About our study

- Longitudinal Survey (Swedish Labour Force Survey)
- 8 waves
- Find (cost reducing) data collection strategy for the 8th wave
- Data from the survey in January 2016
 - 21 521 sample units in total
 - 2 465 in wave 8

Aim: Reduce cost

2465 sample units

2014			2015				2016
April	July	October	January	April	July	October	January
Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8

Data collection history

- Based on paradata
- Wave 1-7
- How "easy"/"difficult" will it be to collect data in wave 8?

Data collection history, Wave 1-7

SAMPLE UNITS IN WAVE 8		2465
Interview in all waves		946
Interview in 7th wave, nonrespondent in at least one previous wave		429
Interview in 7th wave, noncontact before that		134
Refusal in 7th wave, have been interviewed before		113
Refusal in 7th wave, have never been interviewed before		46
Noncontact in 7th wave, interview in 6th wave		93
Noncontact in all previous waves		144
Noncontact in 7th and 6th wave, have been interviewed before		213
Other		347

Stop calls to these sample units



Aim: Not increase bias

Response rate after
wave 7

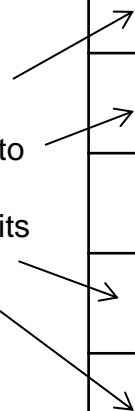
2014			2015				2016
April	July	October	January	April	July	October	January
Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8

Data collection history, wave 1-7

Data collection history, Wave 1-7

SAMPLE UNITS IN WAVE 8		
	<u>Aim of strategy:</u>	
	Reduce cost	Reduce cost + Balance
Refusal in 7 th wave, have been interviewed before	Stop all	Do not stop cases needed for balance
Refusal in 7 th wave, have never been interviewed before	Stop all	
Noncontact in all previous waves	Stop all	
Noncontact in 7 th and 6 th wave, have been interviewed before	Stop all	

Stop calls to these sample units



Proposed Strategies

- Cost reduction (Naive)
 - Paradata
 - Data collection history
- Balanced
 - Low propensity groups after wave 7
 - Auxiliary data
 - Age group
 - Employed or not (according to registers)
 - High education or not (according to registers)

Summary strategies

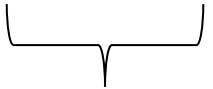
- Cost reduction only
 - Stop sample units unlikely to respond in wave 8
 - Based on history (wave 1-7)
 - Risk to increase bias
- Cost reduction + Balance
 - As above
 - But do not stop sample units that belong to low propensity groups

Evaluation of strategies

- We "apply" the strategies in "retrospect"
- What happens if we remove cases according to strategies?
- How much data is lost?
- How many contact attempts are saved?

Evaluation in retrospect

2014			2015				2016
April	July	October	January	April	July	October	January
Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8



- Outcome in wave 8:
- Interviews
 - Contact attempts

Evaluation: Indicators

P = the weighted response rate in per cent

IMB = the imbalance measure measures the difference between the response set r and the selected sample s for a chosen \mathbf{x} -vector. It could be demonstrated* that IMB is equal to the variance for the response propensities for the chosen \mathbf{x} -vector

$$CV_s = \frac{\sqrt{IMB}}{P}$$

Using *income* from the Tax-register, available for the selected sample s , is it possible to estimate the difference between estimators based on the response set r and the selected sample s .

RDF_{exp} = the relative difference between an expansion estimator and the HT-estimator

RDF_{cal} = the relative difference between a calibration estimator and the HT-estimator

*Särndal & Lundquist 2014

Note: auxiliary variables (register data) depends on available variables and the indicators depends on the sample s .

Evaluation: Results, Wave 8

Strategy	P	m	IMB	CV_s	Income		"Cost /Savings"	
					RDF_{exp}	RDF_{cal}	Interviews lost	Attempts "saved" %
Final	67,7	1670	1,67	19,1	10,5	3,36	0	0
Naïve	65,2	1608	1,98	21,5	12,0	3,95	64	27,5
Balanced 60%	66,2	1632	1,71	19,7	10,9	3,49	38	16,6
Balanced 65%	66,7	1644	1,62	19,1	10,3	3,44	26	13,1
Balanced 70%	67,1	1653	1,64	19,1	10,4	3,47	17	8,0

The response rate P is weighted in percent, m is the number of responding units, IMB , CV_s , RDF_{exp} and RDF_{cal} are multiplied with 100.

Evaluation: Results, Wave 1-8

Strategy	P	m	IMB	CV_s	RDF_{exp}	RDF_{cal}	Interviews “lost”	Attempts “saved”
Final	59,1	12726	1,49	20,6	11,6	4,31	0	0%
Naïve 4	58,8	12664	1,51	20,9	11,8	4,40	64	2,8%
Balanced 4 65%	59,0	12700	1,48	20,6	11,6	4,32	26	1,4%

The response rate P is weighted in percent, m is the number of responding units, IMB , CV_s , RDF_{exp} and RDF_{cal} are multiplied with 100.

Small differences!

Next steps

- Apply stopping rules earlier
 - After wave 6,5,4, etc.
- Models to estimate response propensities
- Effects on survey estimates
- Experiment to evaluate strategies

Thank you

anton.johansson@scb.se

peter.lundquist@scb.se