Minimizing errors in multi-mode surveys through Adaptive Survey Designs

Joep Burger, Nino Mushkudiani,
Koen Perryck and Barry Schouten

Statistics Netherlands, Utrecht University

International Total Survey Error Conference, 21 September 2015, Baltimore

Survey data collection

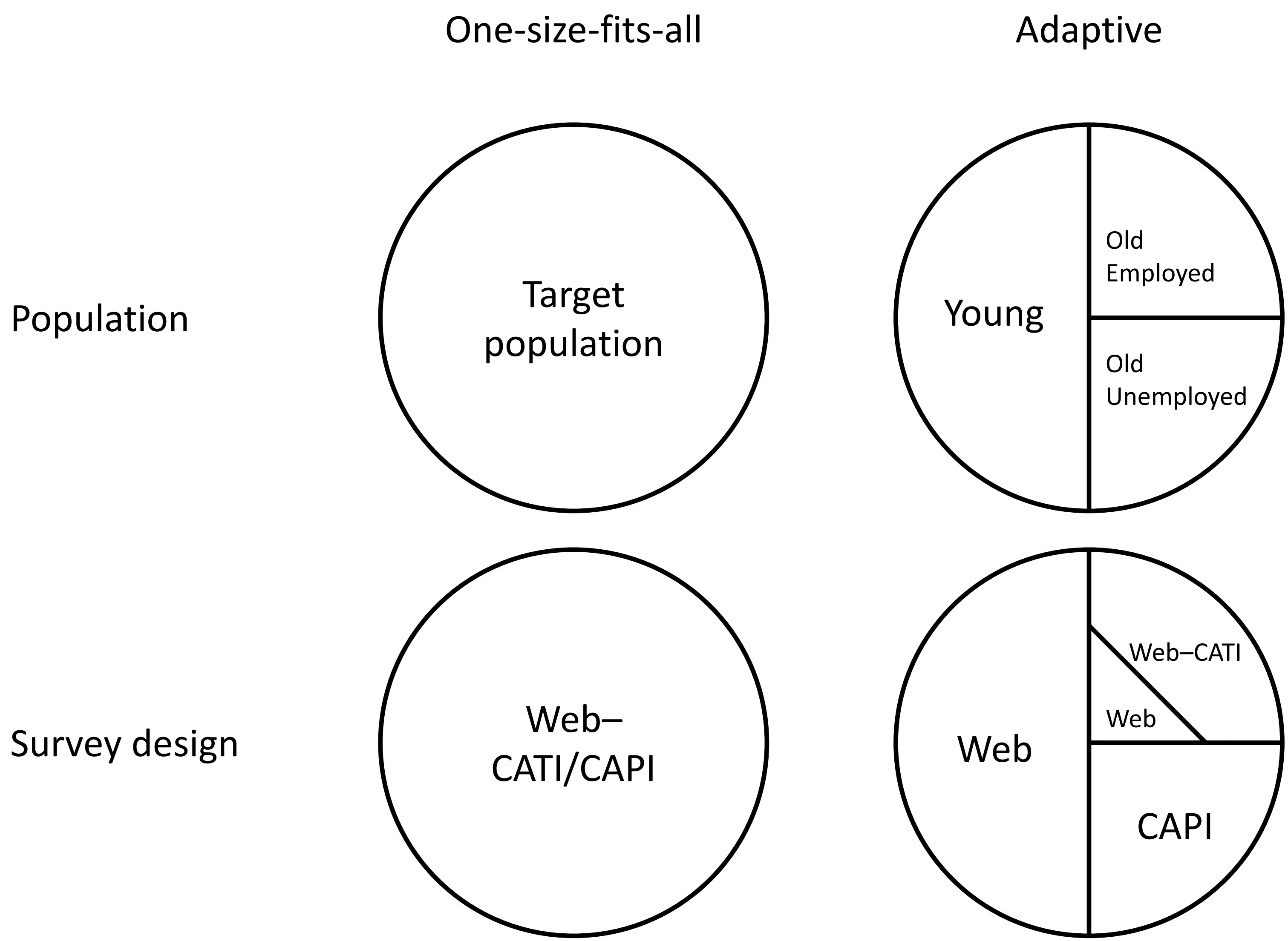
- Trends
  - Declining response rates
  - Budget cuts
- Multi-mode (e.g. web–CATI/CAPI)
  - Cheaper ☺
  - Mode effects ☹
- Accuracy under pressure

Mode effects

- Selection
  - Coverage
  - Nonresponse
- Measurement

Problem if mode mix changes over time

→ Adaptive survey design (ASD)


One-size-fits-all
Adaptive
Population
Target population
Young
Old Employed
Old Unemployed
Survey design
Web–CATI/CAPI
Web
Web–CATI
Web
CAPI

Adaptive Survey Design

- Avoid mode-specific effects
- Maintain budget
- Assign different strategies to different groups
  - Exploit difference between groups

Optimization problem

- Minimize overall mode effect, $\bar{D}$
- By optimally assigning strategies s to groups g

$$\min_{p_{s,g}} \bar{D} = \left| \sum_{s,g} \frac{N_g}{N} \frac{p_{s,g}\rho_{s,g}}{\sum_{s'} p_{s',g}\rho_{s',g}} D_{s,g} \right|$$

Constraints

- Budget

$$\sum_{s,g} N_g p_{s,g} c_{s,g} \leq B$$

- Minimum number of respondents per group

$$r_g = N_g \sum_{s} p_{s,g} \rho_{s,g} \geq R_g$$

- Minimum response rate

$$\frac{r}{n} = \frac{\sum_g r_g}{\sum_{s,g} N_g p_{s,g}} \geq \Gamma$$

Robustness

- ASDs require strategy- and group-specific input parameters
  - Response propensity, $\rho_{s,g}$
  - Sampling cost, $c_{s,g}$
  - Mode effect, $D_{s,g}$
- How sensitive are ASDs to
  - uncertainty about parameter estimates?
  - temporal changes in parameters?

Quantifying robustness

- Compare designs between bootstrap samples
- Compare designs between years
- Effect of old design on subsequent years

Comparing designs

Three distance measures

- relative difference in sample size; $[-1, \infty)$

$$d_1(1,2) = \frac{n_2 - n_1}{n_1}$$

- normalized Euclidean distance between group sampling probabilities; [0,1]

$$d_2(1,2) = \sqrt{\frac{1}{2}\sum_g \left(w_{1,g} - w_{2,g}\right)^2}$$

$$w_{j,g} = \mathbb{P}\{\text{group}|\text{sample}\} = \frac{w_g\left(1-p_{j,\emptyset,g}\right)}{\sum_g w_g\left(1-p_{j,\emptyset,g}\right)};\ w_g = \frac{N_g}{N}$$

- probability that a unit sampled under both designs is not allocated to the same strategy; [0,1]

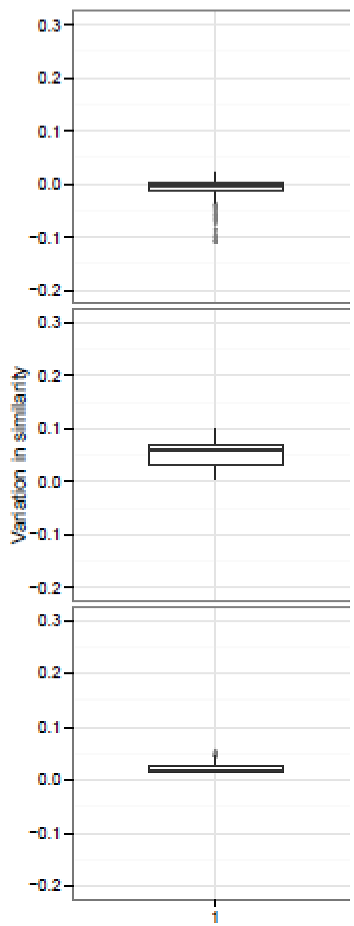
$$d_3(1,2) = 1 - \mathbb{P}\{\text{same strategy}|\text{group}, \text{sample}\}$$
$$= 1 - \sum_{s,g} \frac{w_g\left(1 - p_{1,\emptyset,g}\right)\left(1 - p_{2,\emptyset,g}\right)}{\sum_g w_g\left(1 - p_{1,\emptyset,g}\right)\left(1 - p_{2,\emptyset,g}\right)} P_{1,s,g} P_{2,s,g}$$

$$P_{j,s,g} = \mathbb{P}\{\text{strategy}|\text{group}, \text{sample}\} = \frac{p_{j,s,g}}{1 - p_{j,\emptyset,g}}$$

Case study

- Dutch Travel Survey
  - Travel behavior
- 8 groups: see paper
- 3 strategies
  - Web
  - Web–CATI
  - Web–CATI/CAPI (regular design)
- Parameter estimates: see paper

Robustness against uncertainty



d_1 (sample size)

d_2 (group sampling probabilities)

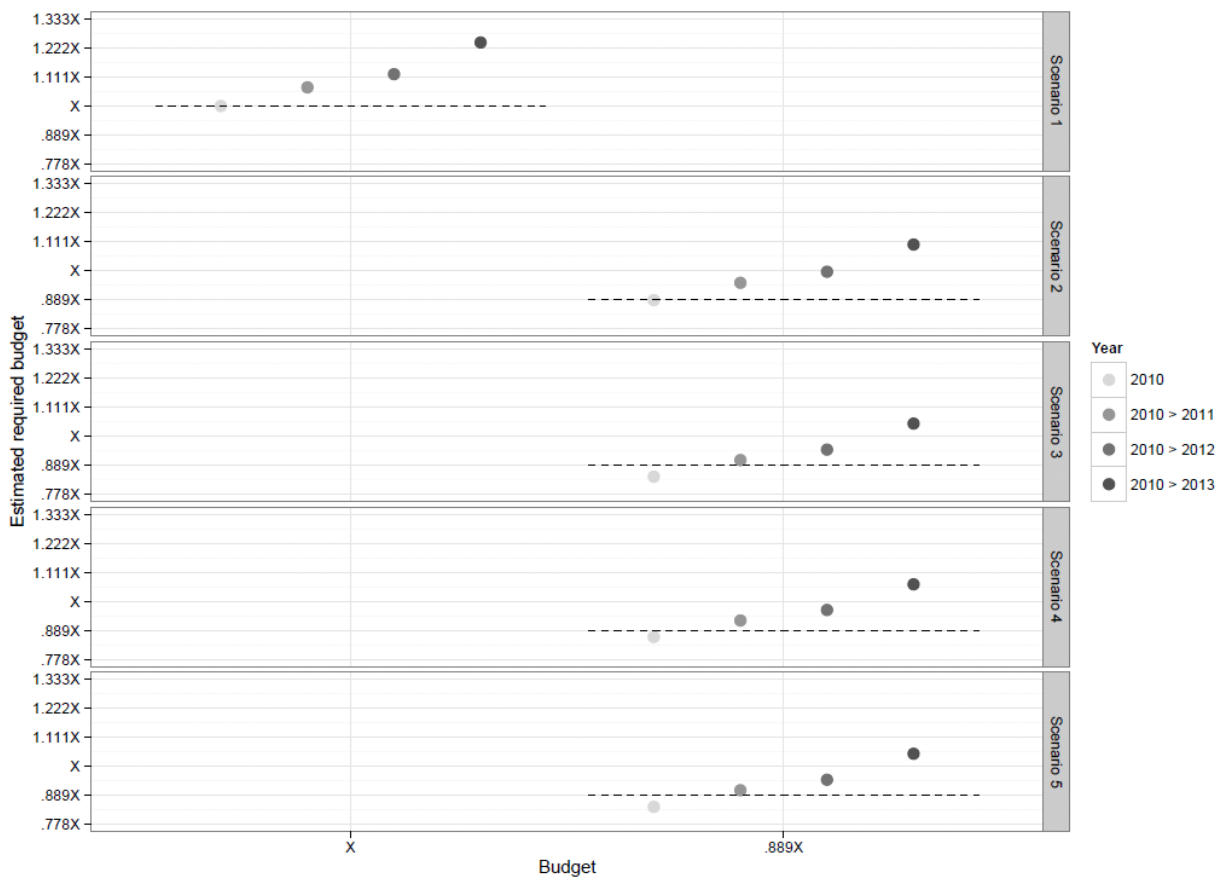
d_3 (conditional allocation probabilities)

Robustness against dynamics

Measure	Scenario	2010–2011	2010–2012	2010–2013
d_1	1	−0.098		
	2			
	3	−0.019		
	4	0.005		
	5	−0.019	0.027	
d_2	1	0.041		
	2			
	3	0.007		
	4	0.013		
	5	0.007		
d_3	1	0.087		
	2			
	3	0.018		
	4	0.158		
	5	0.018	0.145	

No solutions

Continue ASD 2010



Summary

- ASDs can avoid bias-inducing mode effects
  – in contrast to one-size-fits-all MMDs
- Proposed
  – ways to measure robustness
  – measures to compare designs
- Case study
  – robust against uncertainty
  – sensitive to dynamics