

Case prioritization in a survey: Comparing different selection criteria

Tobias Gummer
tobias.gummer@gesis.org

GESIS, Mannheim

4th Workshop “Advances in adaptive and responsive survey design”
11/10/2015, Manchester

Case Prioritization (i)

- Goal: reduction of nonresponse bias
 - Estimate-specific: $bias(Y)$
 - Result of varying response propensities across strata of the sample:
 - $Bias(Y) = \frac{\sigma_{yp}}{\bar{p}}$
- Balance response propensities across strata of the sample
 - Estimate individuals' response propensities
 - Treat low propensity case in order to convert them to respondents
 - "Treatment": devote special attention during fieldwork

Case Prioritization (ii)

Which cases should be prioritized?

Where to set the threshold to define “low propensity cases”?

- Peytchev et al. (2010): 50%
- Rosen et al. (2014): 25%
- Blumenstiel and Gummer (2015): 30%

GESIS-Panel (i)

- Offline-recruited probability-based panel survey
- Mixed mode: mail and web
- Recruitment stage in 2013
- Regular panel waves since 2014
- N=4,488 completed welcome interviews in 2013
- N=4,221 participated in first wave of 2014

GESIS-Panel (ii)

- Panel attrition as a subform of nonresponse
 - Respondents' characteristics available from previous waves:
 - Presumably better explanation model for nonresponse
 - Assess nonresponse bias by comparing variables from welcome interview for whole sample and participants of subsequent waves
- ⇒ “Real” data to build a simulation study

Nonresponse bias

- Difference between the distributions of Y with and without attrition
- Dissimilarity indices as measure of inequality between two distributions (Duncan & Duncan, 1957):

$$D(Y) = 100 \times .5 \times \sum_1^C |f_c^S - f_c^P|$$

- Here:

Difference between Y in the welcome interview for the whole sample and the participants of the first wave of 2014

Table: Example for D.

	w/o attrition	w attrition	$ f_c^S - f_c^P $
female	.5	.6	.10
male	.5	.4	.10
N	4,488	4,221	
D			10.00

Response propensities

- P = participation in first wave of 2014 (0=no,1=yes)
- $\ln(P) = \alpha + \beta_1 X_1 + \dots + \beta_n X_n$
- $Pr(P = 1|X) = \frac{e^{\alpha + \beta_1 X_1 + \dots + \beta_n X_n}}{1 + e^{\alpha + \beta_1 X_1 + \dots + \beta_n X_n}}$
- X : age, gender, education, political interest, civic duty, trust, perceived complexity of the survey, survey experience, willingness to participate, incentive, mode

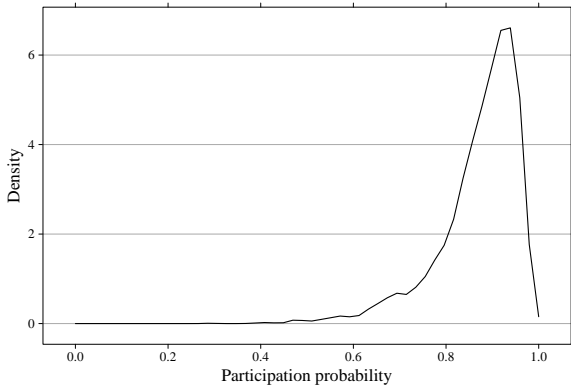


Figure: Density distribution of the predicted probabilities.

Simulation setup (i)

Different selection criteria:

- 1st to 100th percentile of initial response propensities selected ($S=1$) for prioritizing

Recalculate response propensities:

$$\ln(P) = \alpha + \beta_1 X_1 + \dots + \beta_n X_n + tS$$

- t = effect of a treatment
- Three conditions:
 $t = 2 \times \beta_{incentive}, t = 4 \times \beta_{incentive}, t = 6 \times \beta_{incentive}$
- S = case selected to be prioritized (0=no/1=yes)

Resampling of participants in re-interview

- Repeated 500 times

Simulation setup (ii)

Effect of applying case prioritization with a given selection criteria

- Reduction in nonresponse bias

$$= \bar{D}_{prio.} - D_{w/o}$$

- Note: nonresponse bias is estimate-specific and so should be the effect

⇒ Y : age, gender, education, political interest, civic duty, trust, perceived complexity of the survey, survey experience, and willingness to participate

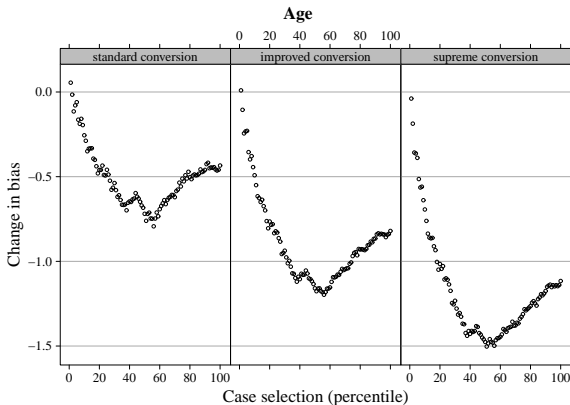


Figure: Reduction of bias due to case prioritization for different selection criteria.

Note: Reduction of bias plotted as difference of D with and without case prioritization.

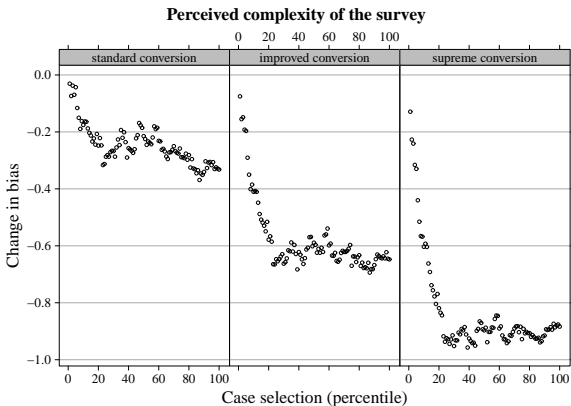


Figure: Reduction of bias due to case prioritization for different selection criteria.

Note: Reduction of bias plotted as difference of D with and without case prioritization.

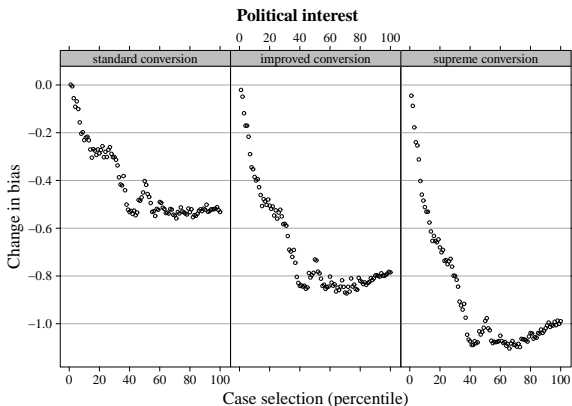


Figure: Reduction of bias due to case prioritization for different selection criteria.

Note: Reduction of bias plotted as difference of D with and without case prioritization.

General findings

- Case prioritization can reduce nonresponse bias
- Selection criteria differ in how effective they are in reducing bias
- There is a minimum nonresponse bias, selecting strata beyond that threshold is less efficient
- Selection criteria appear to be estimate-specific
 - Variation: 20th - 80th percentile

Applicability

- Implementing case prioritization affects the sample irreversibly
- Simulations as pre-study may prove useful
- “Averaged selection criteria” for a set of variables?

Limitations & road ahead

- Cost are an important aspect of efficiency
- Unit nonresponse \neq panel attrition
- Include all variables of welcome interview
- Replicate for participation in other waves
- Improve prediction model
- This is just a case study!

Thank you for listening!

tobias.gummer@gesis.org

Table: Logistic regression on participation in the re-interview.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.1259	0.3417	-3.29	0.0010
Gender: female	0.1509	0.0973	1.55	0.1209
Age	0.0373	0.0036	10.44	0.0000
Education:				
.. low
.. intermediate	0.1694	0.1176	1.44	0.1495
.. high	0.3853	0.1442	2.67	0.0076
Political interest	0.1145	0.0583	1.96	0.0496
Civic duty	-0.0856	0.0602	-1.42	0.1552
Trust	-0.1128	0.0636	-1.77	0.0761
Survey evaluation	0.3181	0.0929	3.42	0.0006
Complexity of survey	-0.4160	0.0863	-4.82	0.0000
Willingness to participate	0.0575	0.0624	0.92	0.3573
Survey Experience	0.2284	0.1401	1.63	0.1030
Incentive	0.1773	0.0959	1.85	0.0645
Mode: onliner	0.8748	0.1027	8.52	0.0000

Note: N=4,418; Nagelkerke $R^2 = 0.11$.

