Can Estimated-Control Calibration Reduce Bias in Estimates from Nonprobability Samples?

Jill A. Dever

Bonnie E. Shook-Sa

Richard Valliant, JPSM

WSS Mini-Conference:
Non-probability Samples
Washington, DC
Sept 9, 2015

Motivation

Fit for Purpose (6 Criteria):

Relevance, Timeliness, Accessibility, Interpretability, Accuracy (Precision), and Coherence [1]

Climate

- Declining response rates
- Measurement errors
- Limited funds
- The need for speed

Motivation

Weights needed for public-/research-use data [2]

- *Propensity score adjustment* [3]
- *Calibration adjustment* [4]

- Composite estimation (multiple data sets)
- Meta-analysis
- Model-based analyses (no weights)
- Bayesian modeling

Two Flavors of Survey Sampling Designs

Probability sampling:

- Presence of a sampling frame linked to population
- Every unit has a known probability of being selected
- Design-based theory focuses on <u>random selection mechanism</u>
- Examples: address-based sampling, dual-frame RDD

Non-probability sampling:

- No population sampling frame available
- Underlying <u>population model</u> is important
- Some opinions on reported estimates of error
- Examples: focus groups, opt-in web panels, quota sampling

Propensity Score Adjustment

Logistic model with a reference survey to estimate probability of selection => weights

Adjust for *selection bias*:

- Covered population
- Catchment area
- Nonresponse (nonparticipation)

Input weights:

- NP weights = 1
- Reference survey weights [4]

Propensity Score Adjustment

Assumptions:

- Surveys are disjoint
- Nonparticipants are missing at random
- Large reference survey from the target population
- Overlap in the questionnaires

Research to date:

- Only part of the bias was removed [5]
- Mixed results [3,6]
- Adjusted reference survey weights needed [4]

Calibration Adjustment

Traditional weight calibration [7]

$$\sum_{s_A} w_k \mathbf{x}_k = \mathbf{t}_{Ux}, \quad \text{where} \;\; \mathbf{t}_{Ux} = \sum_U \mathbf{x}_k$$

$$\hat{t}_{yGR} = \hat{t}_{Ay} + \left(\mathbf{t}_{Ux} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_A$$

Adjust for: [8]

- Coverage
- Nonresponse
- Weight variability

Input weights adjusted for sampling, nonresponse (possibly)

Calibration with Estimated Controls

Estimated Control (EC) Calibration [7]

$$\sum_{s_A} w_k \mathbf{x}_k = \hat{\mathbf{t}}_{Bx}, \quad \text{where } \hat{\mathbf{t}}_{Bx} = \sum_{s_B} w_l \mathbf{x}_l$$

$$\hat{t}_{yEC} = \hat{t}_{Ay} + \left(\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_A$$

Adjust for:

- Coverage
- Nonresponse

Input weights adjusted for sampling, nonresponse (possibly)

Simulation Study

Research questions:

Can estimated-control calibration reduce bias in estimates from non-probability samples?

Is there a difference between EC PSA, PSA.avg and calibration?

Simulation parameters [4]

- 2003 Michigan Behavioral Risk Factor Surveillance Survey (enhanced), N = 50,000
- Volunteer sample selected via Poisson sampling with defined probabilities of participation, n_A = (250, 500, 1000)
- Reference sample selected via simple random sampling, n_B = (1000, 500, 250)
- R = 10,000

Simulation Study

Covariates	Propensity to Volunteer	PSA's	EC Calibration
Age (6)	✓	✓	✓
Race (3)	✓	✓	✓
Gender (2)	✓	✓	✓
Wireless phone (2)	✓	✓	✓
Education (4)	✓	✓	✓
Income (5)	✓	✓	✓
Diabetes (2)		✓	✓

Simulation Study — Result Highlights

Compare relative differences in relative bias with and without estimated control (Diabetes):

$$relbias\left(\hat{\theta}\right) = 100\left(\bar{\hat{\theta}} - \theta\right)\Big/\theta$$

Propensity to Volunteer covariates:

- EC PSA = bias decrease for some (< 5%),
  linked to correlation and size of reference
- EC PSA.avg = higher returns on bias reduction (<5%),
  more volatile results than EC PSA

Simulation Study — Result Highlights

Health Variables (9 categorical, 2 continuous):

- EC PSA = bias decrease for a few and not for others,
  better when reference survey is larger than NP
- EC PSA.avg = higher returns on bias reduction,
  more volatile results than EC PSA
- EC Calibration = bigger bang for the buck,
  more volatile results than EC PSA

Simulation Study — Result Highlights

Relative Sample Size	**Body Mass Index** (*Pct Relative Difference*)		
	PSA	PSA.Avg	Calibration
0.25	0.0	0.0	25.0
1	1.3	0.0	22.7
4	2.5	2.0	19.4

* Reference divided by non-probability sample size

Questions for Future Research

- How “best” to use EC Calibration with Propensity Scores?
- What is the impact on measures of error in using estimated controls?
- How sensitive are the theoretical assumptions underlying the methodology (e.g., surveys must be disjoint)?
- What flavor of estimated control should one choose?

References

[1] Dever, J. A. & Valliant R. (2014). Estimation with non-probability surveys and the question of external validity. Paper presented at the Statistics Canada's *2014 International Methodology Symposium* and published in the forthcoming conference proceedings.

[2] Baker, et al. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, *1*:90–143.

[3] Lee, S. and Valliant, R. (2009). Estimation for volunteer panel Web surveys using propensity score adjustment and calibration adjustment. *Sociological Methods & Research*, *37*(3): 319-343.

[4] Valliant, R., & Dever, J. A. (2011). Estimating propensity adjustments for volunteer web surveys. *Sociological Methods & Research*, *40*(1):105–137.

[5] Tourangeau, R., Conrad, F.G. and Couper M.P. (2013), The Science of Web Surveys, New York: Oxford University Press.

[6] Yeager, et al. (2011), Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly*, *75*(4): 709–747.

[7] Dever, J. A. & Valliant, R. (2010). A comparison of variance estimators for poststratification to estimated control totals. *Survey Methodology*, *36*(1): 45–56.

[8] Kott, P.S. (2006). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors. *Survey Methodology*, *32*(2): 133–142.

More Information

Jill A. Dever

Senior Survey Statistician

Washington, DC

jdever@rti.org