Comparing Alternatives for Estimation from Nonprobability Samples

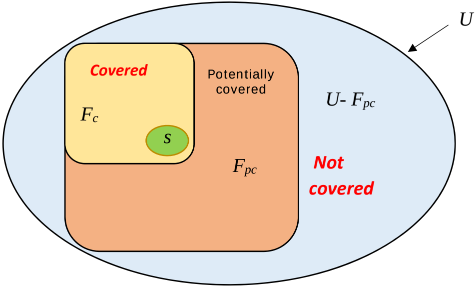
Richard Valliant

Universities of Michigan and Maryland

FCSM 2018

Goals of Weighting and Estimation

Relationship of sample, frame, and target population



- U = target population
- F_{pc} = potentially covered; F_c = actually covered
- $U - F_{pc}$ = not covered at all
- s = sample

Goals of weighting and estimation

- Project sample s to target population U
- Correct for undercoverage by frame, $U - F_{pc}$ (eligible units that cannot be selected)
- Correct for overcoverage by frame (ineligible units that can be selected)
- Undercoverage can be huge problem in nonprobability samples—missing demographic groups (age, race/ethnicity, education levels)

Weighting Nonprobability Samples

Literature

General review [Vehovar, Toepoel & Steinmetz 2016]

Mathematical background
[Elliott & Valliant 2017, Valliant, Dorfman, & Royall, 2000]

AAPOR panel on nonprobability sampling [Baker. et al. 2013]

Evaluation of election polls [AAPOR 2017, Sturgis 2016]

Pew studies
[Kohut, et al. 2012, Kennedy, et al. 2016, Mercer, et al. 2018]

Xbox projection for 2012 US presidential election
[Wang, et al. 2015]

Software examples
[Valliant & Dever, 2018, Valliant, Dever, & Kreuter, 2018]

Approaches to inference

- Quasi-randomization
  - Estimate pseudo-inclusion probabilities and use inverses as weights
- Superpopulation modeling
  - Can give weights that apply to any y if generally useful set of covariates used
- Combine quasi-randomization and superpopulation model
  - Called “doubly robust” in observational data literature

Quasi-randomization with a reference survey

- Reference survey: a probability survey or a census
- Combine reference sample and nonprob sample
- Fit weighted binary regression to predict probability of being in nonprob sample
  - Code nonprob cases = 1, reference cases = 0
  - Wts for nonprob cases = 1, wts for reference cases = survey weight
  - Estimates: Pr(in nonprob sample) within whatever population the reference sample represents
- Weights are inverses of "pseudo-inclusion" probabilities
- Justification is like repeated sampling in design-based world

Superpopulation modeling

- Reference survey unneeded
- Fit linear regression model of y on covariates
- Use fitted model to predict values for nonsample cases
- Add sample values to nonsample predictions to estimate pop total
  - Estimated total is approximately $\hat{t} = \mathbf{t}_{Ux}^{T}\widehat{\beta}$
  - Predict for every unit in population and add up
  - Only pop totals of x's are needed—not individual x's for nonsample units
- Justification: estimator of total is model-unbiased if model is correct

Standard error estimation

- Quasi-randomization: use design-based variance estimator for with-replacement sampling
  - Ignores fact that pseudo-probabilities are estimates
  - Could replicate to reflect that (jackknife, bootstrap)
- Superpopulation modeling
  - Model-based variance estimators are available
  - Replication also works
- Combination (doubly-robust)
  - Need to replicate to reflect all sources of variability

Multilevel regression & poststratification

- Variation on superpopulation modeling
- Fit an elaborate model for a poststratum of units
- Estimate a mean or proportion as

$$\hat{\bar{y}} = \sum_{\gamma=1}^{G} \hat{P}_\gamma \hat{\mu}_\gamma$$

$\hat{P}_\gamma$ = estimated proportion of pop in poststratum γ
$\hat{\mu}_\gamma$ = estimated mean per element in poststratum γ

- PS mean is estimated by random (or mixed) effects model or Bayesian modeling approach
- Begin with cross-classification of many covariates and dynamically decide which crosses to retain

Simulation Study

Simulation population

- Pop based on Michigan BRFSS dataset in R `PracTools` pkg [Valliant, Dever, & Kreuter 2017]
- Demographics—age, race, education, income
- Analysis vars—Self-reported general health, smoked 100 cigarettes in lifetime, good or better health, mean health rating
- $N = 50,000$ in U; select samples from persons who have Internet access at home (about 65% of pop)
- Similar to [Valliant & Dever, 2011] study but samples here are selected to have substantial undercoverage problem

Table: Distribution by age of population of persons, subgroup that has Internet access at home, and samples

	Proportions		
Age	Population	Internet	Sample
1 = 18-24	0.06	0.06	0.12
2 = 25-34	0.14	0.16	0.31
3 = 35-44	0.20	0.23	0.19
4 = 45-54	0.22	0.25	0.20
5 = 55-64	0.17	0.18	0.13
6 = 65+	0.22	0.12	0.05
Total	1.00	1.00	1.00

Table: Proportions of analysis variables in full population and subpop of persons with Internet access at home.

	Smoked 100 cigarettes		Good or better health	
	Population	Internet	Population	Internet
Yes	0.54	0.50	0.84	0.89
No	0.46	0.50	0.16	0.11

Internet subpop smokes less and is more healthy $\Rightarrow$ weighting has a lot to correct for

Sampling & Estimation

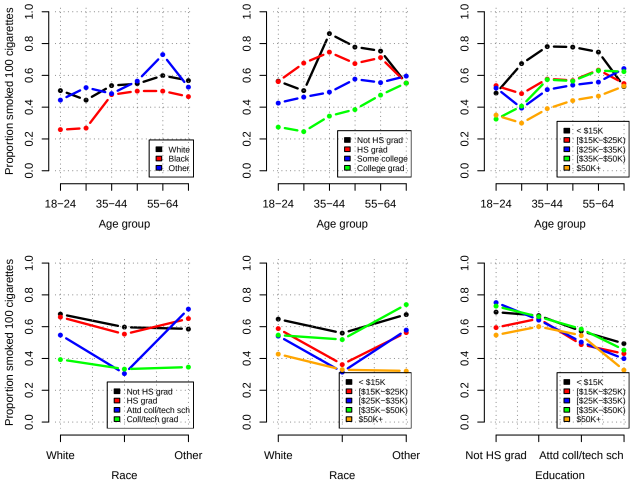


Figure: Proportions in target pop who have smoked at least 100 cigarettes in lifetime by 2-way crosses of age, race, education, and income

Sampling & Estimation

- Select stratified *srswor* sample from Internet subpop; strata = age
  Fact that sample is *stsrswor* is unknown to analyst
- Sample sizes $n = 500$ and $1,000$
- Reference sample is same size as the nonprob sample
- Repeat 10,000 times
- Estimators
  - Quasi-random
  - Model-based: M1 (linear), M2 (raking)
  - Doubly robust (quasi-random + M1)
- SE estimates: WR design-based; grouped JK ($G = 50$) with all est steps repeated in each group

Results

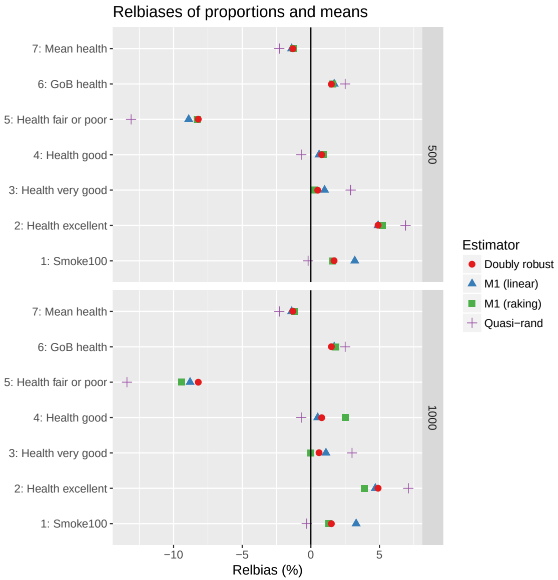


Figure: Quasi-rand has largest |relbias| (with a few exceptions)

Results

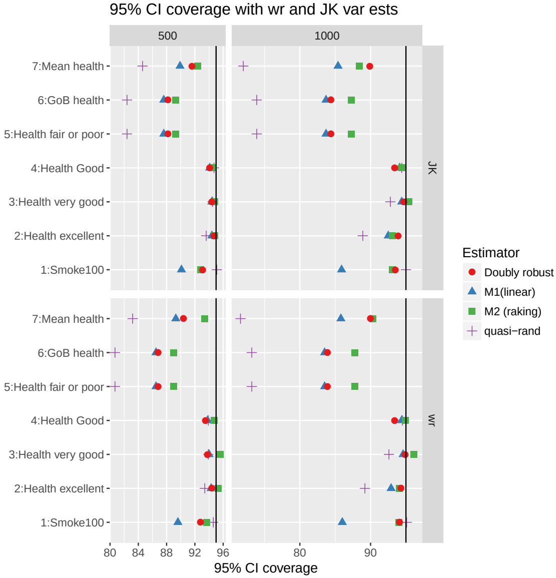


Figure: CI coverage is worse for larger n when estimates are biased

Conclusion

- Poor population coverage is difficult to overcome
- Quasi-random was least effective in eliminating bias
- Linear model & raking have similar relbiases
- Jackknife was somewhat better than WR replacement variance estimator

Conclusion (continued)

Caveat: Everything we do is model-based one way of another

- Nonresponse adjustment—depends on explicit or implicit adjustment model
- Calibration—efficiency depends on fit of model used to calibrate
- Coverage error correction—done either through NR adjustment or calibration
- Quasi-randomization—inclusion model
- Superpopulation—structural model

Conclusion (continued)

Caveat: Everything we do is model-based one way of another

- Nonresponse adjustment—depends on explicit or implicit adjustment model
- Calibration—efficiency depends on fit of model used to calibrate
- Coverage error correction—done either through NR adjustment or calibration
- Quasi-randomization—inclusion model
- Superpopulation—structural model

Conclusion (continued)

Caveat: Everything we do is model-based one way of another

- Nonresponse adjustment—depends on explicit or implicit adjustment model
- Calibration—efficiency depends on fit of model used to calibrate
- Coverage error correction—done either through NR adjustment or calibration
- Quasi-randomization—inclusion model
- Superpopulation—structural model

Conclusion (continued)

Caveat: Everything we do is model-based one way of another

- Nonresponse adjustment—depends on explicit or implicit adjustment model
- Calibration—efficiency depends on fit of model used to calibrate
- Coverage error correction—done either through NR adjustment or calibration
- Quasi-randomization—inclusion model
- Superpopulation—structural model

Conclusion (continued)

Caveat: Everything we do is model-based one way of another

- Nonresponse adjustment—depends on explicit or implicit adjustment model
- Calibration—efficiency depends on fit of model used to calibrate
- Coverage error correction—done either through NR adjustment or calibration
- Quasi-randomization—inclusion model
- Superpopulation—structural model

Conclusion (continued)

Caveat: Everything we do is model-based one way of another

- Nonresponse adjustment—depends on explicit or implicit adjustment model
- Calibration—efficiency depends on fit of model used to calibrate
- Coverage error correction—done either through NR adjustment or calibration
- Quasi-randomization—inclusion model
- Superpopulation—structural model

Conclusion (continued)

Caveat: Everything we do is model-based one way of another

- Nonresponse adjustment—depends on explicit or implicit adjustment model
- Calibration—efficiency depends on fit of model used to calibrate
- Coverage error correction—done either through NR adjustment or calibration
- Quasi-randomization—inclusion model
- Superpopulation—structural model

References I

- Valliant R & Dever JA (2018).
  Survey Weights: A Step-by-step Guide to Calculation. College Station: StataPress.
- Valliant R, Dever, JA, & Kreuter, F (2018).
  Practical Tools for Designing and Weighting Sample Surveys, 2nd edition. New York: Springer.
- Valliant R, Dorfman A & Royall RM (2000).
  Finite Population Sampling and Inference: A Prediction Approach. New York: Wiley.

References II

- Baker R, Brick M, Bates N, Couper M, Courtright M, Dennis J, et al.(2010). AAPOR report on online panels. *Public Opinion Quarterly*, 74, 711–781.
- Elliott, MR & Valliant, R (2017). Inference for Nonprobability Samples. *Statistical Science*, 32(2), 249–264.
- Valliant R & Dever JA (2011). Estimating propensity adjustments for volunteer web surveys. *Sociological Methods and Research*, 40: 105-137.
- Valliant R, Dever JA, & Kreuter F (2017). PracTools: Tools for Designing and Weighting Survey Samples. R package version 0.6. http://CRAN.R-project.org/package=PracTools.
- Vehovar V, Toepoel V, & Steinmetz S (2016). Non-probability sampling. in *The SAGE Handbook of Survey Methodology*, chap. 22. London: Sage.
- Wang W, Rothschild D, Goel S, & Gelman A (2015). Forecasting Elections with Non-representative Polls. *International Journal of Forecasting*, 31, 980–991.

References III

- Baker R, Brick JM, Bates NA, Battaglia M, Couper MP, Dever JA, Gile K, & Tourangeau, R (2013). *Report of the AAPOR Task Force on Non-Probability Sampling*. The American Association for Public Opinion Research. http://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/NPS_TF_Report_Final_7_revised_FNL_6_22_13.pdf
- Kennedy C, Blumenthal M, Clement S, Clinton J, Durand C, Franklin C, McGeeney K, Miringoff L, Olson K, Rivers D, Saad L, Witt E, and Wiezien, C. (2017). *An Evaluation of 2016 Election Polls in the U.S. Ad Hoc Committee on 2016 Election Polling*. The American Association for Public Opinion Research. http://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx
- Kennedy C, Mercer A, Keeter S, Hatley N, McGeeney K & Gimenez A (2016). Evaluating Online Nonprobability Surveys. Pew Research Center report, 2 May 2016. http://www.pewresearch.org/2016/05/02/evaluating-online-nonprobability-surveys/

References IV

- Kohut A, Keeter S, Doherty C, Dimock M, & Christian L (2012). Assessing the representativeness of public opinion surveys.
  http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys
- Mercer, A. Lau, A, & Kennedy, C (2018). For Weighting Online Opt-In Samples, What Matters Most?.
  http://www.pewresearch.org/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/
- Sturgis, P, Baker, N, Callegaro, M, Fisher, S, Green, J, Jennings, W, Kuha, J, Lauderdale, B, and Smith, P (2016). *Report of the Inquiry into the 2015 British general election opinion polls*
  http://eprints.ncrm.ac.uk/3789/1/Report_final_revised.pdf