Non-probability Sampling for Finite Population Inference

Jill A Dever & Richard Valliant

RTI International & Universities of Michigan and Maryland

AAPOR Webinar (October 18, 2016)

Webinar Goals

- Understand the different types of non-probability samples currently in use
- Understand how non-probability samples can be affected by errors such as coverage and nonresponse
- Understand what methods of estimation can be used for non-probability samples and the arguments used to justify them

Motivation for Non-probability Sampling

- Low response rates for many probability samples (Kohut et al. 2012)
- Ever increasing costs with ever decreasing funds
- Nonsampling errors
- The need for speed
- Data are everywhere just waiting to be analyzed!!!

New Sources, New Problems

Examples of “New-ish” Sources of Data

- Twitter
- Facebook
- Snapchat
- Mechanical Turk
- SurveyMonkey
- Web-scraping
- Pop-up Surveys
- Data warehouses
- Probabilistic matching of multiple sources

New Sources of Data: Example Studies

- Analysis of medical records including text to predict heart disease (Giles & Wilcox 2011)
- Correlates of local climate & temperature with spread of infectious disease (Global Pandemic Initiative)
- MIT's Billion Prices Project–Price indexes for 22 countries from web-scraped data
- Marketing of e-cigarettes (Kim et al. 2015)
- Political polls and political issues (e.g., Clement 2016; Conway et al. 2015; Dropp & Nyhan 2016)
- Prediction of social stability (e.g., Kleinman 2014)
- Public health events, outbreaks (Harris et al. 2014; Kim et al. 2012)
- Research on subscribers to PatientsLikeMe.com
- Ad-hoc surveys via Amazon's Mechanical Turk
- Google flu and dengue fever trends (defunct)

Probability vs. Non-probability Samples

Probability sampling:

- Presence of a sampling frame linked to population
- Every unit has a known probability of being selected
- Design-based theory focuses on random selection mechanism
- Probability samples became touchstone in surveys after Neyman (*JRSS* 1934) article

Non-probability sampling:

- Investigator does not randomly pick units with *KNOWN* probabilities
- No population sampling frame available/desired
- Underlying population model is important
- Differing opinions on reporting estimates of error

Probability vs. Non-probability Samples

- We focus on surveys with the goal to use sample to make estimates for *entire finite population*—external validity
- Many applications of big data analysis use non-probability samples. Population may not be well defined.
- Many probability surveys have such low RRs they basically are non-probability samples
  • Pew Research response rates in typical telephone surveys dropped from 36% in 1997 to 9% in 2012 (Kohut et al. 2012)
- Recommendations for using non-probability samples:
  - AAPOR task force reports on non-probability samples (2013) & online samples (2010)
  - Perils and potentials of self-selected entry (Keiding & Louis 2016)

Three Categories of Non-probability Samples

- **Convenience**—units at hand selected; notion overlaps with accidental, availability, opportunity, haphazard or unrestricted sampling
- **Matched**—units are drawn into study (panel) based on characteristics, i.e., controlled selection
- **Network**—a set of units form starting seeds, which sequentially lead to additional units selected (aka snowball, respondent driven sampling)

  (*Note*: Sirken network sampling is an exception)

Types of Convenience Samples

- **Volunteer sampling**—recruitment at events (e.g. sports, music, etc.) and other locations (e.g. mall intercept, street recruitment), limited (if any) refusal conversion
- **River sampling**—general or study-specific invitation through banner/pop-up web ads, etc.
- **Mail-in surveys**—type of volunteer sampling with paper-and-pencil questionnaires, distributed as leaflets at public locations (e.g. hotels, restaurants) or enclosed in magazines, journals, newspapers, etc.)
- **Tele-voting** (text message)—type of volunteer sampling where people are invited to express their vote by calling-in or by sending a text (TV shows, contests)
- **Observational**—"you get what you see"

Types of Matched Samples

- **Purposive sampling—**selection follows some judgment or arbitrary ideas of looking for a "representative" sample
- **Expert selection—**subject experts pick the units, e.g., two most typical settlements selected from a region
- **Quota sampling—**sample "improved" by obtaining targeted socio-demographic quotas (e.g. region, gender, age) to reflect population distribution
- **Balanced sampling**

Comments on Balanced Sampling

- Samples selected until means or other quantities match the population (Särndal et al. 2003)
- Estimates are either unweighted (e.g., *average*) or via a model
- Quota sampling is a subset and focuses only on observable characteristics
- Shown to protect against misspecified inferential models (Royall & Herson 1973; Valliant et al. 2000)
- For probability-based balanced sampling
  - Survey weights are required (e.g., Horvitz-Thompson estimation)
  - Cube method randomly chooses from a set of balanced samples (Deville & Tillé 2004)

Survey Errors

- Coverage
- Selection bias
  - ▶ Coverage and/or selection bias is a problem if the seen (sample) part of the population differs from the unseen (nonsample) in such a way that the sample cannot be projected to the full population
- Nonresponse
  - ▶ (some unknown nonresponse for non-probability surveys)
- Attrition
- Measurement error (e.g., satisficing—provide an acceptable answer instead of the correct one)

Non-probability Electoral Polls: Many Failures

Early failure of a non-probability sample

- 1936 Literary Digest mail survey
- 2.3 million subscribers plus automobile and telephone owners
- Predicted landslide win by Alf Landon over FDR
- Excluded core lower-income supporters of FDR

More recent failures

- British parliamentary election May 2015
  Sturgis et al. (2016) is a post mortem
- Israeli Knesset election March 2015
- Scottish independence referendum, Sep 2014
- US state of Michigan democratic primary, 2016

Non-probability Electoral Polls: One that Worked

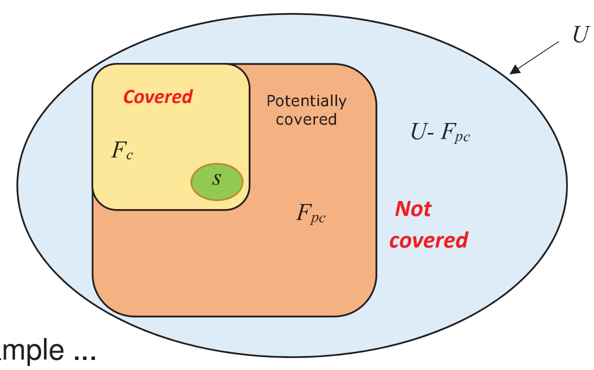
- Xbox gamers: 345,000 people surveyed in opt-in poll for 45 days continuously before 2012 US presidential election
- Xboxers much different from overall electorate:
  - 18- to 29-year olds were 65% of dataset, compared to 19% in national exit poll
  - 93% male vs. 47% in electorate
- Unadjusted data suggested landslide for Romney
- Wang et al. (2015) used multilevel regression and poststratification to get good estimates with covariates
  - sex, race, age, education, state, party ID, political ideology, and who they voted for in the 2008 presidential election

Comparing Probability and Non-probability Samples

Mixed results

- Kennedy et al. (2016)—compared 9 non-probability and 1 probability sample
- Dutwin & Buskirk (2016)—some techniques show benefits (e.g., sample matching) but ....
- Tourangeau et al. (2013)—examined wt adjustments for 8 opt-in web panels using weight with mixed results
- Yeager et al. (2011)—compared RDD and non-probability internet survey with results varying by type of variable
- Valliant & Dever (2011)—effective propensity scores are possible with weighted reference survey cases

Universe & Sample



For example ...

- U = adult population
- $U - F$ = adults without internet access
- F_{pc} = adults with internet access
- F_c = adults with internet access who visit some webpage(s)
- s = adults who volunteer for a panel

Illustration of a Coverage Problem

- Volunteer web panel surveyed about voting intentions
- Support for 2 candidates differs by age group
- Suppose the panel has no one in older groups

A	B	C	D	F	G	H
	Percentage of voters favoring (fictitious)			Proportion of total Presidential	B*F	C*F
Age group	Candidate A	Candidate B		vote in 2012	Candidate A	Candidate B
18 to 24	47	30		0.09	4	3
25 to 44	46	35		0.30	14	11
45 to 64	40	42		0.39	16	16
65 to 74 years	30	50		0.13	4	6
75+	25	70		0.09	2	7
Total				1.00	40	43
Total excluding 65 and older					43	38
Total excluding 75 and older					41	40

Correcting for Sample Imbalance

- Quota sampling or other type of controlled recruiting (YouGov/Polymetrix); no weights needed
- Weights to correct imbalance of sample compared to pop

- Two approaches to weighting
  1. Quasi-randomization weighting
  2. Superpopulation modeling of y's

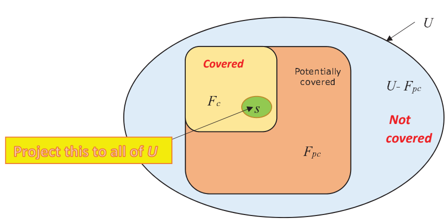
Both involve modeling

Flavors of Missing Data

- **MCAR** (Missing completely at random)
  —Every unit has same probability of appearing in sample
- **MAR** (Missing at random)
  —Probability of appearing depends on covariates known for sample and nonsample cases
- **NMAR** (Not missing at random)
  —Probability of appearing depends on covariates and y's

Population Inference: Estimating a Total

- Pop total $t = \sum_s y_i + \sum_{F_c - s} y_i + \sum_{F_{pc} - F_c} y_i + \sum_{U-F} y_i$
- To estimate t, predict 2nd, 3rd, and 4th sums



- What if non-covered units are much different from covered?
- Difference from a bad probability sample with a good frame but low RR:
  - No unit in $U - F$ or $F_{pc} - F_c$ had any chance of appearing in the sample

Population Inference: Quasi-randomization Approach

Model probability of appearing in sample

$$Pr(i \in s) = Pr(has\ Internet) \times$$
$$Pr(visits\ webpage \mid Internet) \times$$
$$Pr(volunteers\ for\ panel \mid Internet,\ visits\ webpage) \times$$
$$Pr(participates\ in\ survey \mid Internet,\ visits\ webpage,\ volunteers)$$

Probabilities are sometimes estimated with special *Reference* (probability) sample or an existing sample (ACS, NHIS, etc.)

Population Inference: Quasi-randomization Approach

Propensity score method:

- Put s and reference sample together
- Estimate pseudo-inclusion probability $\pi_i = Pr(i \in s)$ via binary regression
- Use $1/\pi_i$ as a weight
- Model covariates:
  - demographic items
  - webographic (attitudinal) items
    –mixed results (Schonlau et al. 2007; Lee et al. 2009))
  - covariates highly correlated with y's (Lee 2006; Dever et al. 2015)

Population Inference: Quasi-randomization Approach

Binary regression to estimate propensity scores:

- Code non-probability cases = 1; reference cases = 0
- $w_{non-prob} = 1$ for non-probability sample cases
- w_{Ref} = survey weight for reference survey cases
- Propensities estimate probability of being in non-prob sample within whatever pop the reference weights to. Cases:
  - $w_{Ref} \uparrow$ adult pop with internet access
  - $w_{Ref} \uparrow$ adult pop regardless of internet access
- Caveats—reference survey weighting must correct for any coverage and nonresponse error
- Poststratification, raking, or other calibration often applied after getting pseudo-inclusion probabilities

Population Inference: Quasi-randomization Approach

Assumptions important for propensity score methods (Valliant & Dever 2011):

- Surveys are disjoint (no respondent overlap)
- Nonparticipants in non-probability survey are MAR
- Large reference survey from target population
- Identical key items on covariates in both questionnaires
- Propensity scores:
  - have common support in reference and non-probability (distributions overlap)
  - estimated with reference survey weights

Population Inference: Superpopulation “Prediction” Approach

- Use a model to predict the value for each nonsample unit (Valliant et al. 2000)
- Linear model: $y_i = \mathbf{x}_i^T \beta + \epsilon_i$
- If this model holds, then

$$
\begin{aligned}
\hat{t} &= \sum_s y_i + \sum_{F_c - s} \hat{y}_i + \sum_{F_{pc} - F_c} \hat{y}_i + \sum_{U - F_{pc}} \hat{y}_i \\
&= \sum_s y_i + \mathbf{t}_{(U-s),x}^T \widehat{\beta} \\
&\doteq \mathbf{t}_{Ux}^T \widehat{\beta}; \qquad \hat{y}_i = \mathbf{x}_i^T \widehat{\beta}
\end{aligned}
$$

Note: Nonlinear models require individual x’s for nonsample units

Population Inference: Superpopulation (Prediction) Approach

$\hat{t} = \sum_s y_i + \mathbf{t}_{(U-s),x}^T \widehat{\beta}$

- $\widehat{\beta} = \mathbf{A}_s^{-1} \mathbf{X}_s^T \mathbf{y}_s$, where $\mathbf{A}_s = \mathbf{X}_s^T \mathbf{X}_s$
- $\mathbf{X}_s$ is $n \times p$ matrix of covariates for the sample units
- $\mathbf{y}_s$ is the n-vector of sample y’s

Resulting weight:

$$w_i = 1 + \mathbf{t}_{(U-s),x}^T \mathbf{A}_s^{-1} \mathbf{x}_i$$

where $\mathbf{t}_{(U-s),x}$ = vector of x totals for nonsample units

Note: With this $\widehat{\beta}$, weights do not depend on y’s
Similar structure to generalized regression estimation (GREG)

y's & Covariates

- If y is binary, a linear model is being used to predict a 0-1 variable
  - Done routinely in surveys without thinking explicitly about a model
- Every y may have a different model $\Rightarrow$ pick a set of x's good for many y's
  - Same thinking as done for GREG and other calibration estimators
- Undercoverage: use x's associated with coverage
  - Also done routinely in surveys

Modeling Considerations

- Good modeling should consider how to predict y's and how to correct for coverage errors
- Covariate selection: LASSO, CART, random forest, boosting, other machine learning methods
- Covariates: an extensive set of covariates needed
  (Dever, Rafferty & Valliant 2008; Valliant & Dever 2011; Wang et al. 2015)
- Model fit with sample needs to hold for nonsample (difficult [impossible?] to prove)

Pros and Cons with Quasi-Randomization and Superpopulation

Quasi-randomization

- Pro = general weights for estimating any y
- Con = possible bias with respect to the superpopulation model for π_i

Superpopulation

- Pro = model-specific estimators with lower variance than quasi-randomization
- Con = possible bias with respect to the superpopulation model for y_i

Notes: Model misspecification a worry for both

Bayesian variations available for each

See review paper by Elliott & Valliant (forthcoming)

Software

Quasi-randomization

- Propensity classes: `pclass` function in R `PracTools` package (Valliant et al. 2015)
- `WTADJUST` and `WTADJX` in SUDAAN 11 (Kott 2016; RTI 2012)
- Custom-written software in SAS, Stata, R, etc.

Superpopulation modeling

- `calibrate` function in R `survey` package (Lumley 2014)
- ReGenesees in R (Zardetto 2015)
- `WTADJUST` and `WTADJX` in SUDAAN 11 (Kott 2016; RTI 2012)
- `ipfraking` in Stata (Kolenikov 2014)
- `sreweight` in Stata (Pacifico 2014)
- `svycal` in future version of Stata

  Set weights to 1 in design-based calibration routines

Simulation Study: Set-up (Valliant & Dever 2011)

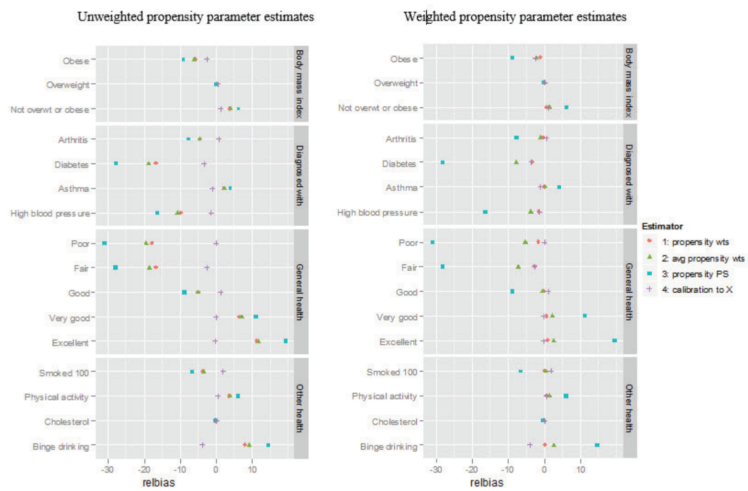
- Data: 2003 Michigan Behavioral Risk Factor Surveillance Survey (MI BRFSS)
- 2,845 sample persons bootstrapped to $N = 50,000$ study population
- $R = 10,000$ simulation runs with two samples:
  1. Volunteer sample
     - Volunteers selected by Poisson sampling; $n = 500$ (expected)
     - Logistic regression for volunteering; probabilities based on having internet access
     - Volunteering probabilities generated with logistic regression with covariates: age, race, gender, wireless phone, education, income
  2. Reference sample—*srswor* of $n = 500$ from non-volunteers

Simulation Study: 4 Estimators Evaluated

1. individual propensity weights (1: propensity wts)
2. average propensity weights in each of five subclasses (2: avg propensity wts)
3. propensity-poststratified estimator (3: propensity PS)
4. calibration to population totals of covariates (no propensity adjustment) using a regression estimator (4: calibration to X); example of a prediction estimator

10,000 simulations with 500 in each volunteer & reference samples

Simulation Study: Statistical Results



Simulation Study: Key Findings

- Reference survey weights need to be used to estimate propensities of volunteering
- Estimates with individual propensities or average propensity weights within classes are biased with unweighted propensity estimates, but less so with weighted
- Propensity-poststratification poor with unweighted or weighted propensity estimates
- GREG and estimate with individual propensity weights generally have smallest biases
- If probability of volunteering depends on y analysis variables, all estimators are biased

Other Research

- Desire to compare estimates from non-probability against "the truth" leads researchers to contrast probability and non-probability surveys
- Quasi-randomization techniques do not always work
  (e.g., Dever & Brown 2016; Willis et al. 2015; Rothschild & Goel 2014; Valliant & Dever 2011; Yeager et al. 2011; Lee & Valliant 2009; Schonlau et al. 2009; Rivers 2007; Duffy et al. 2005)
- Limited comparisons with model-based estimation
- Lingering concerns
  - Were right covariates available?
  - Were they used correctly—multiway interactions?
  - Poor modeling leads to biased estimators

Variance estimation

Quasi-randomization

- Treat pseudo-inclusion probabilities in same way as designed-based selection probabilities
- Design-based variance estimators apply. Justification is consistency under quasi-randomization distribution
- Linearization or replication can be used
  Replication shows most promise (Lee & Valliant 2009)
- Need to decide whether strata and clusters are appropriate

Superpopulation modeling

- Compute variance under model used for point estimates with variance based on squared residuals
- Replication estimators also justified (Valliant et al. 2000)
- Bayesian models, e.g., credibility interval (Santos, Buskirk & Gelman 2012) with(out) applying survey design effects
- Justification is consistency under superpopulation model

Takeaway

- Non-probability samples do not have the (false?) assurance of complete population coverage that probability samples do
- Inference to finite populations is possible but only with either correct modeling of
  - Chance of being in sample, or
  - Dependence of analysis variables on covariates
- Convincing users that a non-probability sample represents nonsample part of population will always be an issue (true for low RR probability samples, too)

Diagnostics

- Work needed on diagnostics for "representativity"
- Are non-probability estimates aiming at desired target population?
- Distance measure
  $\hat{\theta}$ = set of estimates from non-probability sample
  θ = values from some reliable data source (ACS, NHIS, CPS, census, etc.)

$$D^2 = (\hat{\theta} - \theta)^T [cov(\hat{\theta} - \theta)]^{-1} (\hat{\theta} - \theta)$$

  Compare to a chi-square distribution or F
- Validation items in $\hat{\theta}$ are not used in non-probability weight calculation; may not be of direct interest in the survey

The Future .....

- Quasi-randomization—model pseudo-inclusion probabilities
- Superpopulation models—model the y's
- Combination

Which is better???

The Future

The Future .....

- Quasi-randomization—model pseudo-inclusion probabilities
- Superpopulation models—model the y's
- Combination

Which is better???

39 / 43

References

- Baker R, Brick JM, Bates N, Couper M, Courtright M, Dennis J, et al (2010). AAPOR report on online panels. *Public Opinion Quarterly*, 74:711–781.
- Baker R, Brick JM, Bates NA, Battaglia M, Couper MP, Dever JA, Gile K, & Tourangeau R (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, 1:90–143.
- Clement S (2016). How The Washington Post-SurveyMonkey 50-state poll was conducted. The Washington Post, 6 Sep 2016. Accessed 25 Sep 2016.
- Conway BA, Kenski K & Wang D (2015). The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary. Journal of Computer-Mediated Communication, 20(4):363–380.
- Deville J-C & Yves Tillé Y (2004). Efficient balanced sampling: the cube method. *Biometrika*, 91(4):893–912.
- Dever JA & Brown D (2016, May). Estimated-Control calibration for non-probability surveys. Presented at AAPOR 71st annual conference, Austin, TX.
- Dever JA, Rafferty A & Valliant R (2008). Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias? *Survey Research Methods* 2:47–62.
- Dever JA, Shook-Sa BE & Valliant R. (2015, Sep). Can estimated-control calibration reduce bias in estimates from nonprobability samples? Presented at WSS Mini-Conference on Non-probability Samples, Washington, DC.
- Dropp K & Nyhan B (2016). The 2016 Race: It Lives. Birtherism Is Diminished but Far From Dead. The New York Times, 23 Sep 2016. Accessed 25 Sep 2016.
- Duffy B, Smith K, Terhanian G & Bremer J (2005). Comparing data from online and face-to-face surveys. *International Journal of Market Research*, 47(6):615–639.
- Dutwin D & Buskirk TD (2016). Apples to Oranges or Gala vs. Golden Delicious: Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples. Presentation at the AAPOR 71st annual conference, Austin, TX.
- Elliott MR & Valliant R (2017). Inference for non-probability samples. *Statistical Science*, 31(1):forthcoming.
- Giles T & Wilcox R (2011). IBM Watson and Medical Records Text Analytics. Presentation at the Healthcare Information and Management Systems Society meeting. `https://www-01.ibm.com/software/ebusiness/jstart/downloads/MRTAWatsonHIMSS.pdf`. Accessed 25 Sep 2016.

References

- Harris JK, Mansour R, Choucair B, Olson J, Nissen C, & Bhatt J (2014). Health Department Use of Social Media to Identify Foodborne Illness—Chicago, Illinois, 2013–2014. *Morbidity and Mortality Weekly Report*, 63(32):681–685.
- Keiding N & Louis TA (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys. *Journal of the Royal Statistical Society*, Series A, 179(2):319–376.
- Kennedy C, Mercer A, Keeter S, Hatley N, McGeeney K & Gimenez A (2016). *Evaluating Online Nonprobability Surveys*. Pew Research Center report, 2 May 2016. Accessed 1 Oct 2016. `http://www.pewresearch.org/2016/05/02/evaluating-online-nonprobability-surveys/`
- Kim AE, Hansen HM, Murphy J, Richards AK, Duke J & Allen JA (2013). Methodological Considerations in Analyzing Twitter Data. *Journal of the National Cancer Institute Monographs*, 47:140–146.
- Kim AE, Hopper T, Simpson S, Nonnemaker J, Lieberman AJ, Hansen H, Guillory J, & Porter L (2015). Using Twitter Data to Gain Insights into E-cigarette Marketing and Locations of Use: An Infoveillance Study. *J Med Internet Res*, 17(11):e251.
- Kim A, Richards AK, Murphy JJ, et al. (2012). Can automated sentiment analysis of Twitter data replace human coding? Paper presented at American Association for Public Opinion Research Annual Conference; May 18, 2012; Orlando, FL.
- Kleinman A (2014). Facebook Can Predict With Scary Accuracy If Your Relationship Will Last. The Huffington Post, 14 Feb 2014. Accessed 25 Sep 2016.
- 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`, 15 May 2012. Accessed 9 Sep 2016.
- Kolenikov, S (2014) Calibrating survey data using iterative proportional fitting (raking). *Stata Journal* 14(1): 22–59.
- Kott PS (2016). Calibration weighting in survey sampling. *WIREs Computational Statistics*, 8:39–53.
- Lee S (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of Official Statistics*, 22(2):329–349.
- Lee S & Valliant R (2009). Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociological Methods & Research*, 37:319–343.

References

- Lumley, T (2014). survey: analysis of complex survey samples. R package version 3.30. http://CRAN.R-project.org/package=survey.
- Neyman J (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97:558–625.
- Research Triangle Institute (2012). *SUDAAN Language Manual*, Volumes 1 and 2, Release 11. Research Triangle Park, NC: Research Triangle Institute. www.rti.org/sudaan
- Rivers D (2007). Sampling for web surveys. Proceedings of the *American Statistical Association*, Section on Survey Research Methods. Paper presented at the Joint Statistical Meetings, Salt Lake City, UT. http://s3.amazonaws.com/static.texastribune.org/media/documents/Rivers_matching4.pdf
- Rothschild D & Goel S (2014). Non-Representative Surveys: Fast, Cheap and Mostly Accurate. Presentation given at the AAPOR 70st annual conference, Hollywood, FL.
- Royall R & Herson J (1973). Robust estimation in finite populations I. *Journal of the American Statistical Association*, 69:880-889.
- Santos R, Buskirk T, & Gelman A (2012). Understanding a “credibility interval” and how it differs from the “margin of sampling error” in a public opinion poll. Statement from AAPOR Ad Hoc Committee on Credibility Interval, 7 Oct 2012. Accessed 10 Oct 2016. https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/DetailedAAPORstatementoncredibilityintervals.pdf
- Särndal CE, Swensson B, & Wretman J (2003). *Model Assisted Survey Sampling*, Section 14.4. New York, NY: Springer-Verlag, Inc.
- Schonlau M, Van Soest A, & Kapteyn A (2007). Are ‘Webographic’ or attitudinal questions useful for adjusting estimates from Web surveys using propensity scoring? *Survey Research Methods*, 1(3):155-163.
- Schonlau M, van Soest A, Kapteyn A, & Couper M (2009). Selection Bias in Web Surveys and the Use of Propensity Scores. *Sociological Methods Research*, 37(3):291–318.
- Sirken M (1970). Household surveys with multiplicity. *Journal of the American Statistical Association*, 65:257–266.
- 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.

References

- Tourangeau R., Conrad FG, & Couper MP (2013). *The Science of Web Surveys*. New York:Oxford University Press.
- 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 (2015). PracTools: Tools for Designing and Weighting Survey Samples. R package version 0.3. http://CRAN.R-project.org/package=PracTools.
- Valliant R, Dorfman A & Royall R (2000). Finite population sampling and inference: a prediction approach. New York: Wiley.
- Vehovar V, Toepoel V, & Steinmetz S (2016). Non-probability sampling. *The SAGE Handbook of Survey Methodology*, chap. 22. London: Sage.
- Yeager DS, Krosnick JA, Chang L, Javitz HS, Levendusky MS, Simpser A, & Wang R (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly*, 75:709–747.
- Wang W, Rothschild D, Goel S, & Gelman A (2015). Forecasting elections with nonrepresentative polls. *International Journal of Forecasting*, 31:980–991. http://www.stat.columbia.edu/~gelman/research/published/forecasting-with-nonrepresentative-polls.pdf
- Willis GB, Chowdhury SR, de Moor JS, Ekwueme D, Kent E, Liu B, et al. (2015). A Comparison of Surveys Based on Probability Versus Non-Probability Sampling Approaches. Presentation given at the AAPOR 70st annual conference, Hollywood, FL.
- Zardetto, D (2015). ReGenesees: an Advanced R System for Calibration, Estimation and Sampling Error Assessment in Complex Sample Surveys. *Journal of Official Statistics*, 31(2): 177–203.