Survey Weighting

Richard Valliant

Universities of Michigan and Maryland

AAPOR Webinar (November 14, 2017)

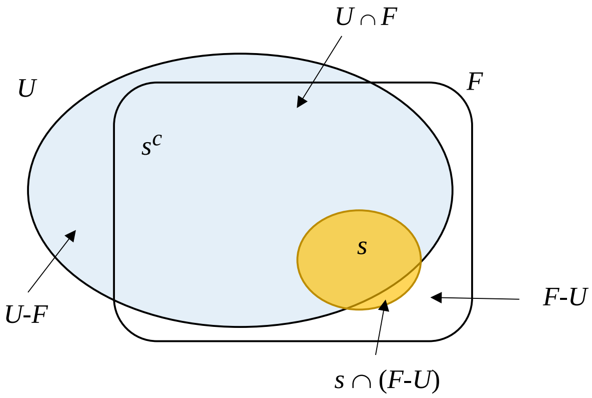
1. Webinar Goals
2. Goals of Weighting and Estimation
3. General Steps in weighting probability samples
   - Base Weights
   - Nonresponse adjustments
   - Using Auxiliary Variables for Calibration
4. Nonprobability samples
   - Quasi-randomization
   - Superpopulation modeling
5. References

Webinar Goals

- Understand how weights are used to correct for coverage errors and nonresponse
- Understand the different steps in weighting and the reasoning behind each
- Understand how weighting approaches differ for probability and nonprobability samples
- Review software for weighting

Goals of Weighting and Estimation

Relationship of sample, frame, and target population



- U = target population
- F = sample frame
- s = sample

Goals of weighting and estimation

- Project sample s to target population U
- Correct for undercoverage by frame, $U - F$ (eligible units that cannot be selected)
- Correct for overcoverage by frame, $F - U$ (ineligible units that can be selected)

Basic properties of weights

- w_i = weight for sample unit i
- Usually $w_i \geq 1$
- $\sum w_i y_i$ estimates pop total of an analysis variable y
  No. of persons who voted for candidate A
  No. of persons unemployed in Nov 2017
  Total amount spent on medical care in 2016
- $\sum w_i$ estimates no. of elements (units) in population
- Estimating pop totals requires weights to be properly scaled

Normalized weights

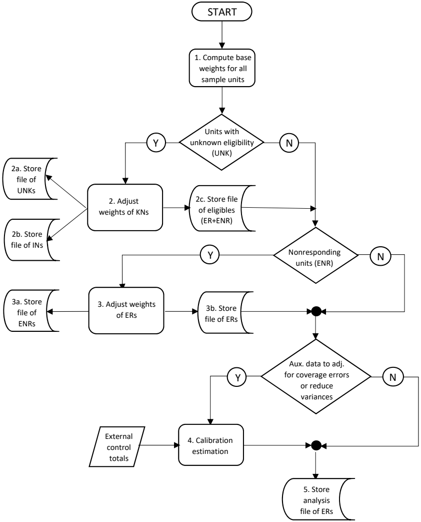
- Adjust weights so that they sum to sample size n of elements
- $w_i^* = n \frac{w_i}{\sum w_i}$
- Normalization has no effect when estimating means or proportions since they have form $\sum w_i y_i / \sum w_i$
- Anachronistic holdover from days when there was no software to analyze survey data
- Makes QC checks to see whether weights are scaled correctly less straightforward

General Steps in Weighting Probability Samples

General steps in probability samples

1. Base weights
2. Unknown eligibility
3. Nonresponse adjustment
4. Calibration to population controls

General steps–schematic picture



Base Weights

Base Weights

- Base weight (aka sampling, design, or selection weight) = inverse of selection probabilities
  $d_{0i} = 1/\pi_i$
- *srs*: $d_{0i} = N/n$
- *stsrs*: $d_{0i} = N_h/n_h$
- Probability proportional to x: $d_{0i} = N\bar{x}_U/nx_i$
- Multistage sample: a weight is computed for each stage of selection and multiplied together to construct the (element) base weight

Two-stage sampling leading to *epsem*

- Select sample of students in two stages—schools at 1st, students at 2nd. PSUs are schools selected with probabilities proportional to size (*pps*) of student body. Equal probability sample of $\bar{n}$ students selected in each PSU.
- School selection probs: $\pi_i = mN_i/N$ for school i
- N_i = number of students in school i
- $N = \sum_{i \in U} N_i$ = total number of students in the population
- Conditional student selection prob is $\pi_{j|i} = \bar{n}/N_i$ for any student j within school i.
- Overall probability of selection is $\pi_{ij} = \pi_i \pi_{j|i} = \frac{mN_i}{N} \frac{\bar{n}}{N_i} = \frac{m\bar{n}}{N}$
- Base weight for student j in school i is $d_{0ij} = \pi_{ij}^{-1} = N/m\bar{n}$. *Self-weighting* since each student has *same* base weight.

Software for sample selection & base weight calculation

1. R package `sampling`
   - `strata` function selects stratified samples (*srswor*, *srswr*, *pps*)
   - `cluster` function
   - `UPsystematic`, `UPrandomsystematic`, and other functions for single-stage
   - Put together for multistage, e.g., `cluster` followed by `strata`
2. Stata—need to do some programming, e.g., `for` loop for stratified sample; some user-written routines
3. SAS—`proc surveyselect` will select many types of samples
4. Write your own

Advantages of commercial or open source software:

- Most bugs are identified if there are many users
- Standard software insures uniform quality across surveys
- Personnel cost savings
- Well-vetted random number generators are used

Nonresponse Adjustments

Methods of analyzing nonresponse

- Deterministic

  Every unit is an R or an NR, no random choice

- Stochastic response

  Every unit has a probability of being an R or an NR

  Also called quasi-randomization

Types of Stochastic Missingness

- Missing Completely at Random (MCAR)–every unit has same probability of response. Respondents are just a random subsample of initial sample.
- Missing at Random (MAR)–probability of response does not depend on y but does depend on some or all of the auxiliaries x. Response model can be formed that depends on x if auxiliaries known for both respondents and nonrespondents.
- Nonignorable nonresponse (NINR) aka Not missing at random (NMAR)–chances of responding depend on one or more analysis variables (y's). Dependence cannot be eliminated by modeling response based on covariates (x's).

Ways of adjusting for NR

Goal: estimate probability of response & use inverse as a weight adjustment

- Form classes of units and make common adjustment within each class
- Estimate individual response propensities and adjust with each

Issues when using adjustment cells

- How to form cells guided by analysis of response patterns, or y's, or both
- Bias under stochastic response model [Kalton & Maligalig 1991]

$$B_R\left(\hat{\bar{y}}_\pi\right) \doteq \frac{1}{N\bar{\phi}} \sum_{i \in U} \left(y_i - \bar{Y}_U\right)\left(\phi_i - \bar{\phi}\right)$$

  $\Rightarrow$ form cells to have common mean of Y or common response propensity ϕ within each cell
- Bias under superpopulation model. [Little & Vartivarian 2005] results say give primacy to forming cells where units all have a common mean.

Issues for Nonresponse Adjustment (continued)

- General approach is to form cells to either
  1. Contain units that all have about same response probability, or
  2. All have a common mean of y
- #2 is hard because there are usually many y's and only means for R's are known.
- #1 is usually more feasible
  - But covariates available for forming cells may be related to both y's and response probabilities
  - Also, a limited number of covariates may be available

Propensity score adjustments

1. Fit model to predict response based on available covariates; $\hat{\phi}(\mathbf{x}_i)$
2. Sort file (R's and NR's both) from low to high based on estimated response propensities
3. Divide into cells (5 to 10 usually enough)
4. Compute NR adjustment in each cell as sum of weights for full sample divided by sum of weights for respondents. Input weights can be base weights or UNK-eligibility adjusted weights for eligible cases.
   Other options: inverses of unweighted RR, mean $\hat{\phi}(\mathbf{x}_i)$; median $\hat{\phi}(\mathbf{x}_i)$
5. Multiply weight of each R in a cell by NR adjustment ratio
6. Only respondents have a non-zero weight after this step.

Software for propensity score adjustments

1. R package `PracTools`; `pclass` function

   ```
   pclass(formula, data, link="logit", numcl=5,
   type, design=NULL)
   ```

   `formula` = binary regression model of form `response~terms`
   `numcl` = no. of propensity classes to form
   `type` = survey-weighted or unweighted regression
   `design` = design object created by `survey` package

2. Stata—programming required; sort file (both R's and NR's) from low to high based on estimated response propensities, use `egen` wih `cut` function

Using Auxiliary Variables for Calibration

Idea behind Use of Auxiliaries in Calibration

- Use relationship between analysis variables (y's) and covariates (x's) to improve estimators
- Reduce variances
- Correct coverage errors
- Need
  - population totals of x's (or good estimates of them) and
  - x's for individual responding sample units
- No need for x's for individual nonsample units and non-responding sample units

Software for calibration

- R `survey` package on CRAN: `postStratify, rake, calibrate` functions
- R `ReGenesees` package: not on CRAN, download from ISTAT
- Stata `svyset` with `poststrata` option
- Stata `svycal` procedure: postratification, raking, general regression estimation (to be released in future version)
- Stata `ipfraking`: raking [Kolenikov 2014]
- Stata `sreweight`: raking, more general calibration [Pacifico 2014]
- SUDAAN `WTADJUST, WTADJX`: [Folsom & Singh 2000, Kott 2006, Chang & Kott 2008, Kott & Chang 2010]

Examples of calibration: Poststratification (PS)

Method

- Put units into groups (age groups, regions, types of business)
- Adjust weights so that estimated counts of units equal control counts

$$\hat{T}_{yPS} = \sum_{\gamma=1}^{G} N_\gamma \hat{\bar{y}}_\gamma = \sum_{\gamma=1}^{G} N_\gamma \left(\hat{t}_{y\gamma} \middle/ \hat{N}_\gamma \right)$$

- $\hat{t}_{y\gamma} = \sum_{s_\gamma} d_i y_i$ is est'd total of y in poststratum γ based on the input weights d_i (usually base or NR-adjusted weights)
- s_γ is set of sample units in poststratum γ
- $\hat{N}_\gamma = \sum_{s_\gamma} d_i$ is est'd pop size of poststratum γ based on input weights
- N_γ is pop count or control total for poststratum γ, and G is the total number of poststrata.

Poststratification (continued)

- Implied final weight for unit i in poststratum γ is

$$w_i = d_i \frac{N_\gamma}{\hat{N}_\gamma}$$

  where $g_i = N_\gamma / \hat{N}_\gamma$ is PS adjustment factor.
- This is also the *g*-weight if we write the final weight as $w_i = d_i g_i$. With that definition of the weight, $\hat{T}_{yPS} = \sum_{i \in s} w_i y_i$, i.e., a weighted sum of the data values.
- Note: $\sum_\gamma w_i = \frac{N_\gamma}{\hat{N}_\gamma} \sum_\gamma d_i = N_\gamma$ (weights are “calibrated”)

Example of when PS is useful

Hispanic and Age related to a y variable $\Longrightarrow$ use Hispanic $\times$ Age PS

Table: Percentages of persons in `nhis.large` population in `PracTools` who reported receiving Medicaid.

	Age group (years)				
Hispanicity	under 18	18–24	25–44	45–64	65+
Hispanic	32.2	10.7	7.6	11.0	27.2
Non-Hisp White	12.6	6.6	3.8	3.1	3.7
Non-Hisp Black and other race/ethnicity	31.3	12.7	8.8	6.4	16.5

Correcting Undercoverage via Poststratification

- Undercoverage is common in HH surveys—CPS underestimates numbers of young African-American and Hispanic males by 20 to 30%
- $\hat{N}_\gamma = \sum_{s_\gamma} d_k$ in poststratum γ based on input weights $<$ census count
- $N_\gamma / \hat{N}_\gamma > 1$ "corrects" for undercoverage

  Still problems if sample is different from nonsample on y's

Poststratification Software

- R `survey` package

  `postStratify(design, strata, population)`

  `design` = survey design object that defines strata, PSUs, and weights
  `strata` = field that identifies poststrata
  `population` = vector of population poststratum counts of elements
  Retrieve weights from a `design` object with `weights(...)`

- Stata

  `svyset [pweight=...], poststrata(...) postweight(...)`

  `pweight` = field with input weights
  `poststrata` = field with poststratum ID
  `postweight` = pop count for poststratum that contains the record

Raking

- Also commonly used
- Marginal pop counts used for 2 or more variables
- Margins can themselves be crosses of variables
  Raking margins could be age group, gender, ethnicity $\times$ education-level
- Method gives individual weights that do not depend on y's

Raking Software

- R `survey` package
  `calibrate(design, formula, calfun="raking", population)`

  `design` = survey design object that defines strata, PSUs, and weights
  `formula` = expression that specifies raking model, e.g. ~age.grp + hispanic
  `population` = vector of population marginal counts of elements
- Stata new procedure `svycal` in future version
  `svycal rake i.age_grp i.hispr [pw=wt], gen(rake_wt) totals(_cons=8000 1.age_grp=6000 2.age_grp=2000 1.hispr=5000 2.hispr=3000)`

  `i.age_grp` = generates age group as factor
  `gen(rake_wt)` = generate raked weights and save as `rake_wt`
  `totals(...)` = pop count marginal counts

General regression estimation (GREG)

- Use linear regression model to predict y values
- Estimator is approximately unbiased and consistent in repeated sampling
- Technique gives weights that do not depend on any particular y but can be used for all y's
- For example, model whether a person will vote for a candidate as a function of age, gender, party affiliation, education level, income, and interactions
  Need pop totals of registered voters for all variables and interactions used
- Can be used to correct for coverage errors

General regression estimation (GREG)

- R `survey`
  `calibrate(design, formula, population, bounds=c(-Inf,Inf), calfun=c("linear"))`
- Stata
  `svycal regress` *`model spec`* `[pw=...], gen(...) totals(...)`

Weighting Nonprobability Samples

Literature

General review [Vehovar, Toepoel & Steinmetz 2016]

Mathematical background [Elliott & Valliant 2017]

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]

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

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 can be 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, references cases = 0
    - Weights for nonprob cases = 1, weights for reference cases = survey weight
    - Estimates Pr(in nonprob sample) within whatever population the reference sample represents
    - Could smooth out estimated probs by grouping (just as in response propensity approach for NR adjustment)
- Weights are inverses of "pseudo-inclusion" probabilities
- Justification is like repeated sampling in design-based world

See [Elliott & Valliant 2017, Valliant & Dever, SMR 2011]

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

See [Valliant, Dorfman, & Royall, 2000]

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

Software for weighting nonprobability samples

- Quasi-randomization: use same routines as for propensity score estimation for NR adjustment
  `pclass` in R `PracTools`
- Superpopulation modeling: use same routines as for calibration in probability samples
  `calibrate` in R `survey` or `svycal` in Stata
  Set initial weights to 1

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
- Software
  `glmer` in R `lme4` package or `rstanarm` R package

More Details

- **Sampling & weighting**
  - Särndal CE, Swensson B, & Wretman J (2003). *Model Assisted Survey Sampling*. New York, NY: Springer-Verlag.
  - Valliant R & Dever JA (2018). *Survey Weights: A Step-by-step Guide to Calculation*. College Station: StataPress.
  - Valliant R, Dever, JA, & Kreuter, F (2013). *Practical Tools for Designing and Weighting Sample Surveys*. New York: Springer.
- Superpopulation modeling (non-Bayesian)
  - Valliant R, Dorfman A & Royall RM (2000). *Finite Population Sampling and Inference: A Prediction Approach*. New York: Wiley.

Downloadable examples

- R code examples from *Practical Tools for Designing and Weighting Survey Samples*: go to `https://jointprogram.umd.edu/` Click About/Faculty and locate me
- Stata code examples: will be available at `http://www.stata-press.com/data/svywt/` after *Survey Weights: A Step-by-step Guide to Calculation* is published

Conclusion

- Methods for weighting probability samples
- Methods for weighting nonprobability samples
- Software for each

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

- Särndal CE, Swensson B, & Wretman J (2003).
  Model Assisted Survey Sampling. New York, NY: Springer-Verlag
- Valliant R & Dever JA (2018).
  Survey Weights: A Step-by-step Guide to Calculation. College Station: StataPress.
- Valliant R, Dever, JA, & Kreuter, F (2013).
  Practical Tools for Designing and Weighting Sample Surveys. 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.

Chang, T & Kott PS (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika*, 95, 3, 555–571.

Elliott, MR & Valliant, R (2017). Inference for Nonprobability Samples. *Statistical Science*, 32(2), 249–264.

Folsom, RE & Singh, AC (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. *Proc. Survey Res. Meth. Sect.*, 598–603. Washington, DC: American Statistical Association.

Kalton, G. & Maligalig, D.S. (1991). A comparison of methods of weighting adjustment for nonresponse. *Proceedings of the US Bureau of the Census Annual Research Conference*, 409–428.
http://babel.hathitrust.org/cgi/pt?id=umn.31951d00357147b;view=1up;seq=417

Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (raking). *Stata Journal* 14(1): 22–59.

References III

- Kott, PS (2006). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors.
  Survey Methodology, 32(2): 133–142.
- Kott, PS & Chang, T (2010). Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse. *Journal of the American Statistical Association*, 105(491), 1265–1275.
- Little, R.J.A. & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology*, 31, 161–168.
- Lumley, T (2017). survey: analysis of complex survey samples. R package version 3.32. `http://CRAN.R-project.org/package=survey.`
- Pacifico, D. (2014). sreweight: A Stata command to reweight survey data to external totals. *Stata Journal* 14(1): 4–21.
- Research Triangle Institute (2012). *SUDAAN Language Manual*, Volumes 1 and 2, Release 11. Research Triangle Park, NC: Research Triangle Institute. `www.rti.org/sudaan`

References IV

- 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 Intervals, 7 Oct 2012. `https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/DetailedAAPORstatementoncredibilityintervals.pdf`
- 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 V

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.
`http://www.istat.it/en/tools/methods-and-it-tools/processing-tools/regenesees`

References VI

- 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/
- 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

References VII

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