The AmeriSpeak® Experience: Methods of Using a National Probability Sample Panel for Studies that Combine Probability and Nonprobability Samples

Vicki Pineau, Ed Mulrow, N. Ganesh, Michael Yang, and J. Michael Dennis

NORC at the University of Chicago

June 10, 2019

NORC at the UNIVERSITY of CHICAGO

Outline

I. Research problem -- Estimation for combined probability and nonprobability samples
II. NORC's AmeriSpeak Panel and low incidence studies
III. Methods for combining probability and nonprobability samples
IV. Comparative analyses for several AmeriSpeak low incidence studies
V. Proposed best practices framework for measuring and reporting total survey error
VI. Future research directions/questions

Research Problem – Estimation for Combined Probability and Nonprobability Samples

Probability Samples

- Since Neyman (1934) probability sampling has been the standard basis for inference from sample to population
  - Well defined target population
  - Presence of a sampling frame linked to the population of interest
  - Sample design where every frame unit has a known and non-zero probability of being selected
  - Design-based estimation theory based on random selection mechanism

Often Need for Fit-for-Purpose Solution

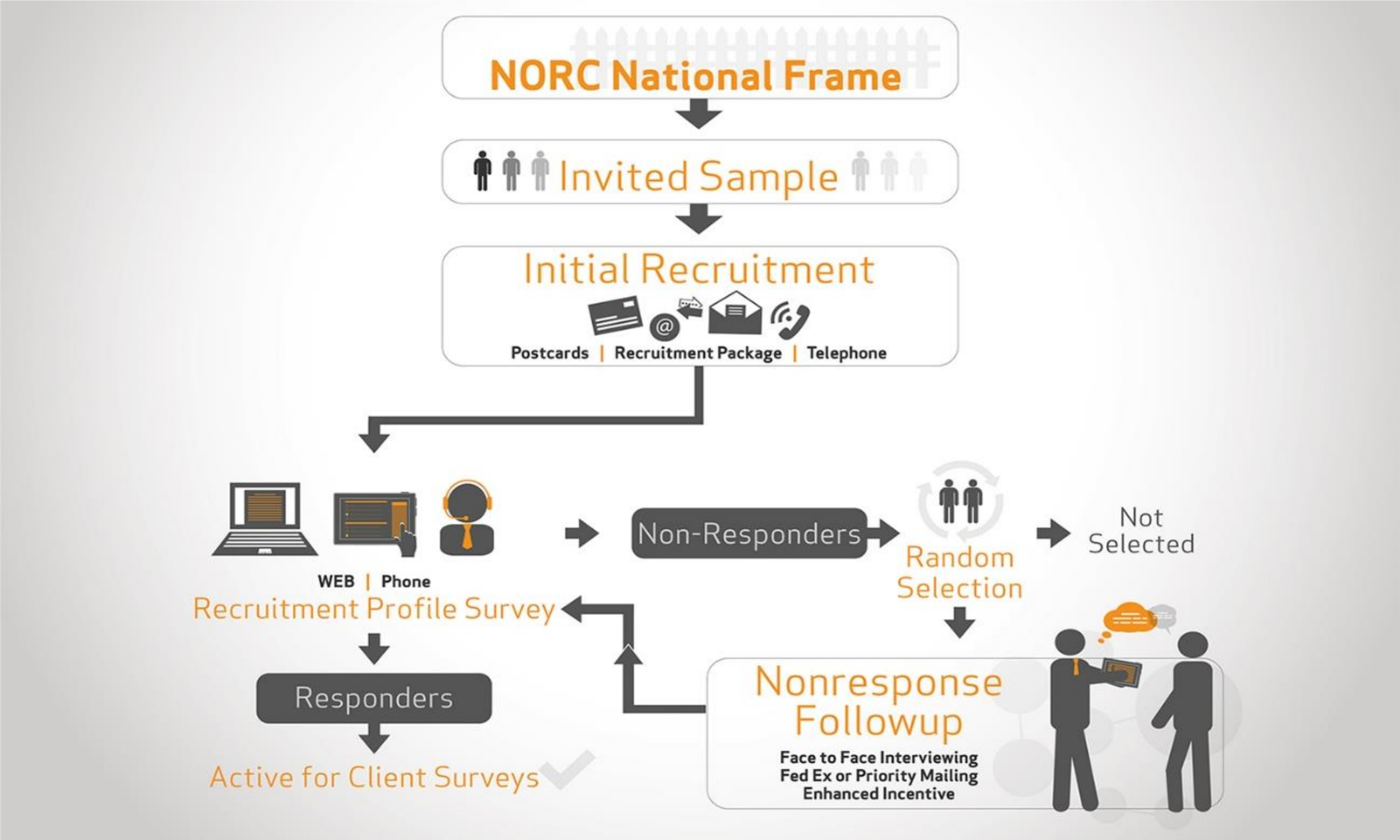
- Studies with an incidence of <10% may require a larger sample than feasible using probability-based methods
- Need a “Fit-for-Purpose” solution to support low incidence studies that combine:
  - Probability-based sample
    - Use as anchor to minimize bias
  - Nonprobability source to decrease cost, supplement to increase sample size, reduce variance, and support small domain estimation
    - Estimation based on combined sample

Total Survey Error Challenges with Nonprobability Samples

- Total Survey Error measurement challenges
  - Lack of scientific design
  - Unknown population coverage
  - Unknown selection mechanism
  - Unknown selection bias
- Design-based inference impossible
  - Model-based inference a necessity (Elliott & Valliant, 2017)

NORC's AmeriSpeak Panel

NORC's AmeriSpeak Panel: Probability-based Sample



NORC at the UNIVERSITY of CHICAGO

AmeriSpeak By the Numbers

Number of Participating Households → (50 States + DC) — **30K**

Client Surveys Completed → (Since June 2015) — **300+**

Panel Recruitment Response Rate → (AAPOR RR3) — **34%**

NORC at the UNIVERSITY of CHICAGO

Methods for Combining Probability and Nonprobability Samples

NORC at the UNIVERSITY of CHICAGO

Five General Methods

1. **Calibration:** Calibrate total estimates to known control totals
2. **Statistical Matching:** Statistically match nonprobability and probability samples
3. **Superpopulation Modeling:** Use a superpopulation model to derive population estimates
4. **Propensity Weighting:** Model the propensity to be included in a nonprobability sample
5. **True North Small Area Modeling**: A superpopulation modeling method developed in-house at NORC (Ganesh et al., 2017), relies on methods commonly used in Small Area Estimation

Method 1: Calibration

- Assign a weight of 1 to all nonprobability sample units.
- Rake the weights to:

1) Known demographic control totals, and
    - **E.g., age, race, gender, income, etc.**
2) Webographic control totals.
    - **“Early adopter” characteristics estimated from NORC’s AmeriSpeak® Panel.**

Method 2: Statistical Matching

- Each nonprobability sample unit is matched to one and only one probability sample unit.
- A match is made by finding a pool of probability sample units with the 20 closest distances to a nonprobability unit, and randomly selecting one unit from the pool.
- Distances are measured using Gower's dissimilarity measure, which allows for both categorical and continuous variables.
- Variables used to calculate distances were identified using Gradient Boosting, an ensemble learning algorithm.
- The nonprobability unit assumes the weight of the matched probability unit.
- When a nonprobability unit is matched to multiple probability units, the nonprobability unit weight is the probability unit weight divided by the number of matches.

Method 3: Superpopulation Modeling

- Assumes that the response variable (*Y*) for the observed sample follows a statistical model
- Different models are fitted for each response variable of interest
- Depending on the type of response variable (e.g., continuous, categorical, count, etc.), different statistical models are used
  - E.g., linear model, logistic regression model, Poisson regression model, etc. See Elliott & Valliant, Valliant et. al., Wang, et. al.
- Variance estimates can be generated under the model
- Auxiliary data are used as predictors in the model
  - Individual level predictors
  - Aggregated geographic, domain predictors
  - Depending on the model, external population totals might be needed for individual level predictors
  - Model selection is required to select the “best” set of covariates

NORC at the UNIVERSITY of CHICAGO

Method 4: Propensity Weighting

1. Concatenate probability sample units and nonprobability sample units, and create a dichotomous variable, *Y*, which is 1 for probability sample units and 0 otherwise.
2. Fit a logistic regression model with *Y* as the response variable.
   - **Covariates include demographic, webographic, and attitudinal/behavioral variables.**
3. Estimate inclusion probabilities for nonprobability sample units from the model.
4. Weights for the nonprobability sample units are the inverse of the predicted inclusion probabilities.

Method 5: True North Small Area Modeling

- Post-stratify weights of prob & nonprob samples independently to population control totals (e.g. age, race, sex, etc.)
  - Probability sample units assigned an appropriate design-based weight
  - Nonprob sample units assigned an input weight of 1.0
- Create weighted survey outcome estimate separately for prob and nonprob samples
  - e.g. Doctor diagnosed allergy; smoking behavior
- Get domain level covariates from external benchmarks (ACS, NHIS) expected to be correlated with survey outcome
  - Domains: e.g. race/ethnicity by age groups; internet use
- Small area modeling to obtain domain level estimates of survey outcomes for combined prob&nonprob samples
- "Small area weights" can be generated by raking the input study weights to the model-based domain level estimates for a key set of response variables

Comparative Analyses

Study 1: Food Allergy Survey with Convenience sample

- Food Allergy Survey data that NORC collected on behalf of Northwestern University.
- Key measures: the adult and child prevalence of self-reported and doctor-diagnosed food allergies, allergy reactions, experiences in allergy treatments, events coinciding with development or outgrowing a food allergy, and perceived risks associated with food allergies. Only the adult data was used for testing.
- Data were collected via both a probability sample and a nonprobability sample.
  - Probability sample: Selected from AmeriSpeak Panel, 7,218 completed surveys.
  - Nonprobability sample: Selected from SSI's opt-in panel, 33,331 completed surveys.

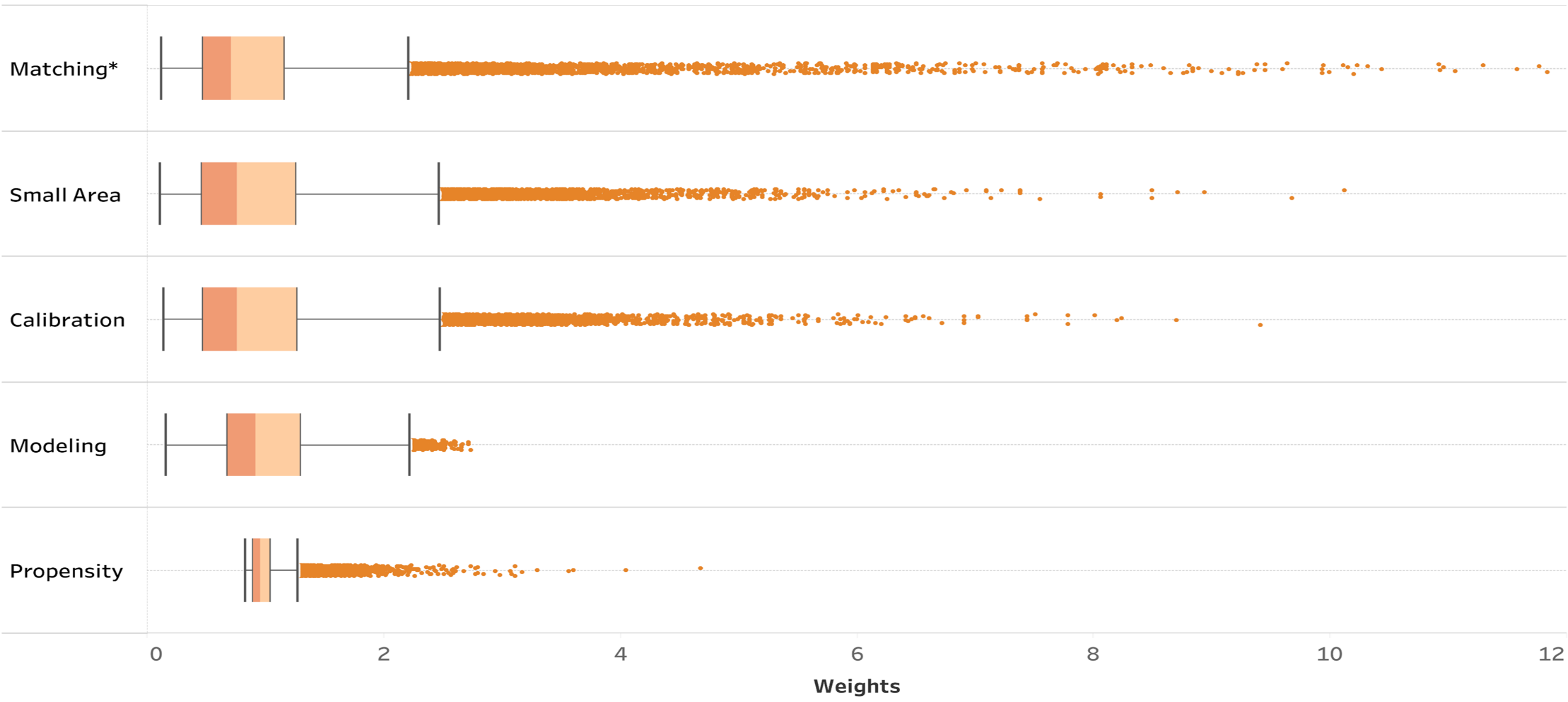
Study 2: Omnibus Survey with Convenience sample

- NORC internal methodological study sample
- A 15-minute survey with wide range of topics, including measurements of political attitudes, views on social issues and the economy, personal finances, participation in social groups, news behavior, personal health and medical care
- Data collected via both a probability sample and a nonprobability sample
  - Probability sample: Selected from AmeriSpeak Panel, 2,681 completed surveys.
  - Nonprobability sample: Selected from SSI’s opt-in panel, 1,243 completed surveys.

Current Focus: Pseudo Weights

- Each method produces a set of pseudo weights for the nonprobability sample units
- The pseudo weights are then used to support weighted estimation
- Each set of weights was scaled so that the sum of the weights equaled the respective nonprobability sample size
- Estimates based on the probability sample are also provided, along with upper and lower 95% confidence bounds
- Reasons for focusing on weights:
  - Not feasible to have a different model for each survey variable
  - Single set of weights provides a robust solution
  - Same thinking as done for GREG and other calibration estimators
  - Clients are used to weighted estimation

Nonprobability Sample Weight Boxplots: Study 1 Food Allergy Data



* Matching had 22 weights great than twelve with a maximum of 28.12

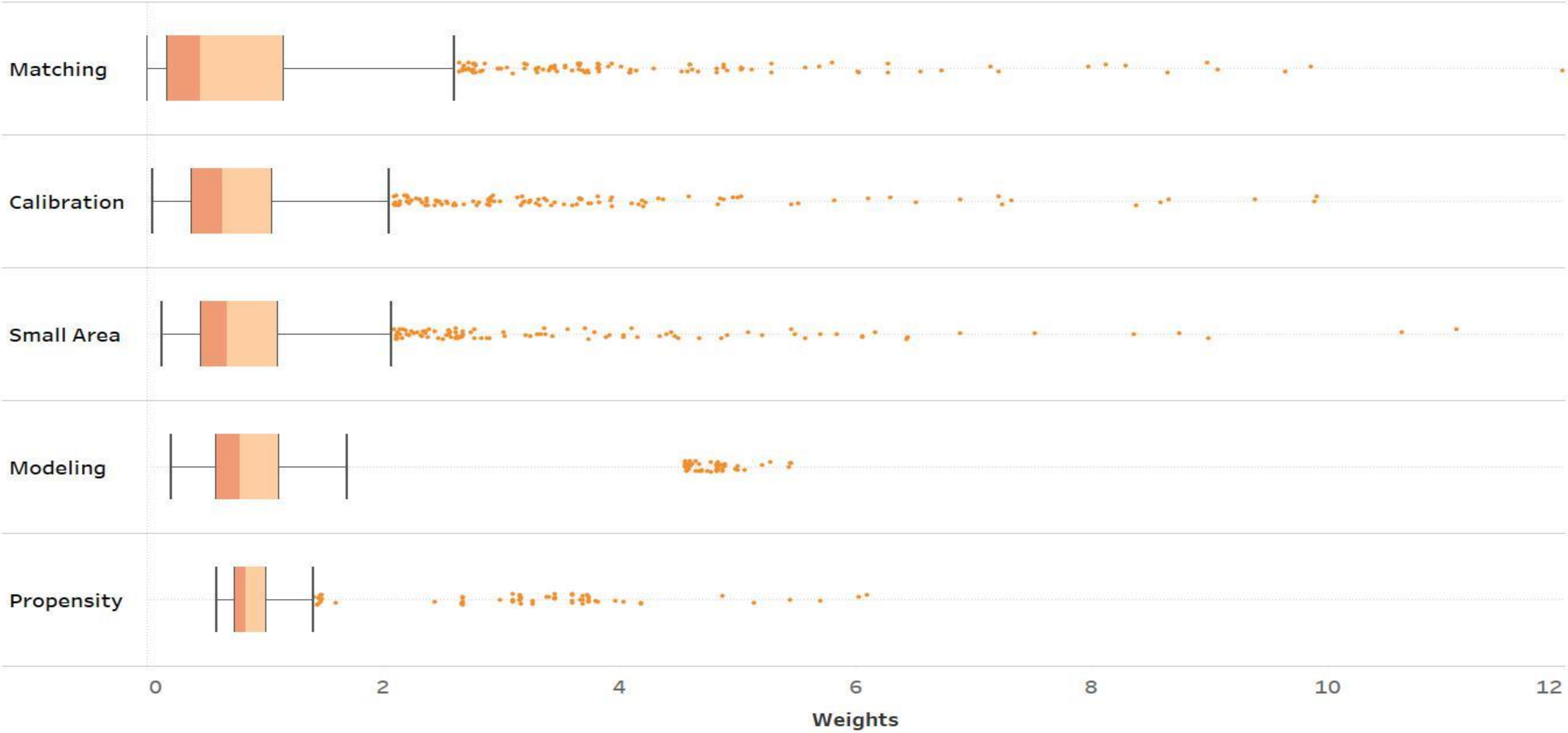
Nonprobability Sample Weighted Estimates: Study 1 Food Allergy Data

Chronic Condition Variable	**Probability Sample Estimates**			**Nonprobability Sample Estimates**				
Have you ever had…	**LCB**	**Mean**	**UCB**	**Calibration**	**Propensity**	**Matching**	**Small Area**	**Modeling**
Doctor diagnosed Asthma	12.0	13.1	14.2	13.1	12.9	13.2	12.0	12.2
Doctor diagnosed Uricaria/chronic hives	0.8	1.0	1.3	1.0	1.0	0.9	0.8	0.8
Doctor diagnosed EoE	0.1	0.2	0.3	0.3	0.3	0.3	0.2	0.2
Doctor diagnosed Diabetes	8.9	9.6	10.4	10.6	10.1	9.7	10.1	10.1
Doctor diagnosed FPIES	0.1	0.2	0.3	0.3	0.3	0.4	0.2	0.3
Doctor diagnosed Eczema	6.7	7.5	8.3	7.0	6.7	6.7	6.4	6.4
Doctor diagnosed Insect sting allergy	3.8	4.4	5.0	4.0	3.9	3.8	3.6	3.5
Doctor diagnosed Latex allergy	2.4	2.8	3.2	2.4	2.6	2.4	2.1	2.2
Doctor diagnosed seasonal allergies	21.8	23.1	24.3	21.9	21.3	21.3	20.6	20.5
A food allergy during your lifetime	20.3	21.6	22.8	28.6	27.8	28.5	21.7	27.5
Doctor diagnosed Medication allergy	14.8	15.8	16.9	13.2	13.2	12.5	12.2	12.5
No doctor diagnosed chronic conditions	46.3	47.9	49.5	51.9	52.5	51.5	52.3	52.4
Doctor diagnosed Other condition	7.9	8.7	9.6	7.1	7.2	6.7	6.7	6.6

Outside confidence bounds

NORC at the UNIVERSITY of CHICAGO

Nonprobability Sample Weight Boxplots: Study 2 Omnibus Survey Data



Nonprobability Sample Weighted Estimates: Study 2 Omnibus Survey Data

Analysis Variable	Probability Sample Estimates			Nonprobability Sample Estimates					Combined Small Area Estimates
	LCB	Mean	UCB	Small Area	Calibration	Matching	Modeling	Propensity	
Voted in 2016 presidential election	69.1	72.3	75.5	72.4	71.4	71.2	73.5	73.2	73.2
Registered to vote	77.9	80.9	83.8	81.3	80.9	79.6	82.4	82.1	82.1
Generally happy	83.9	86.2	88.5	88.8	87.5	82.6	87.5	86.9	86.7
Should be possible to obtain a legal abortion	56.8	60.3	63.7	64.0	66.4	64.1	65.7	65.6	62.4
Country heading in right direction	37.0	40.5	44.0	49.6	45.1	47.2	46.7	47.1	44.0
Most people can be trusted	33.4	36.9	40.3	45.6	44.6	40.1	45.5	44.6	38.9
Nation's economy is good	48.2	51.7	55.1	52.4	46.4	44.0	48.7	47.8	52.4
Household financial situation is good	55.0	58.4	61.7	57.6	51.2	45.8	52.0	51.5	57.6
Government should do more to solve problems	52.9	56.3	59.8	56.1	64.0	63.8	62.7	63.4	56.1
Marijuana should be made legal	60.6	64.0	67.3	67.1	69.0	69.9	67.8	68.9	64.4
Smoked at 100 cigarettes	37.9	41.3	44.7	41.3	46.2	44.4	46.8	45.6	41.3
Should protect the right to own guns	39.9	43.4	46.8	50.6	49.8	53.1	48.1	48.9	45.5

Outside confidence bounds

NORC at the UNIVERSITY of CHICAGO

Key Observations

- No apparent “best” choice of estimation method empirically.
- All methods investigated are model-based and depend on the use of covariates.
- The number and nature of the covariates differ across the methods, which may have contributed to the observed differences in the weights and estimates.
- Methods that rely on explicit models tend to generate less variables weights.
- We have relied mostly on demographic and webographic variables as covariates in implementing the methods. Some important response variables may be weakly correlated with these covariates.
- Because bias may be an issue for nonprobability samples, our intuition suggests that methods that produce larger variances are preferable in the absence of a bias estimate.
- Our preferred method is True North Small Area:
  - It is the only methods that contains explicit bias estimation
  - It generates relatively large variances

Study 3: Smoking Behavior Survey with Respondent Driven Sample (RDS)

- NORC internal methodological study sample
- A 15-minute survey about smoking behavior among 18-55 LGBT population
- Data collected via both a probability sample and a nonprobability sample
  - Probability sample: Selected from AmeriSpeak Panel, 182 completed (seed) surveys.
  - Nonprobability sample: Referred from AmeriSpeak panel completes using Respondnet Driven Sampling, 102 completed (referral) surveys.

	LGBT	Non-LGBT	Total
Seed	182	228	410
Referral	102	0	102
Total	284	228	512

Combining Prob and NonProb Samples: Study 3 Smoking Behavior Data

- Compare estimates from 3 estimation methods, not all 5 previously described:
  - True North small area modeling
  - Propensity weighting
  - RDS estimation -- NEW
    - **Modified Voltz-Heckathorn (V-H) Weighting***
      - Base weight = 1/reported network size
      - Rake base weights to NHIS and CPS population control totals
        - Age group, gender, race/ethnicity

*Gile, Krista J., and Mark S. Handcock. "7. Respondent-Driven Sampling: An Assessment of Current Methodology." *Sociological methodology* 40, no. 1 (2010): 285-327.

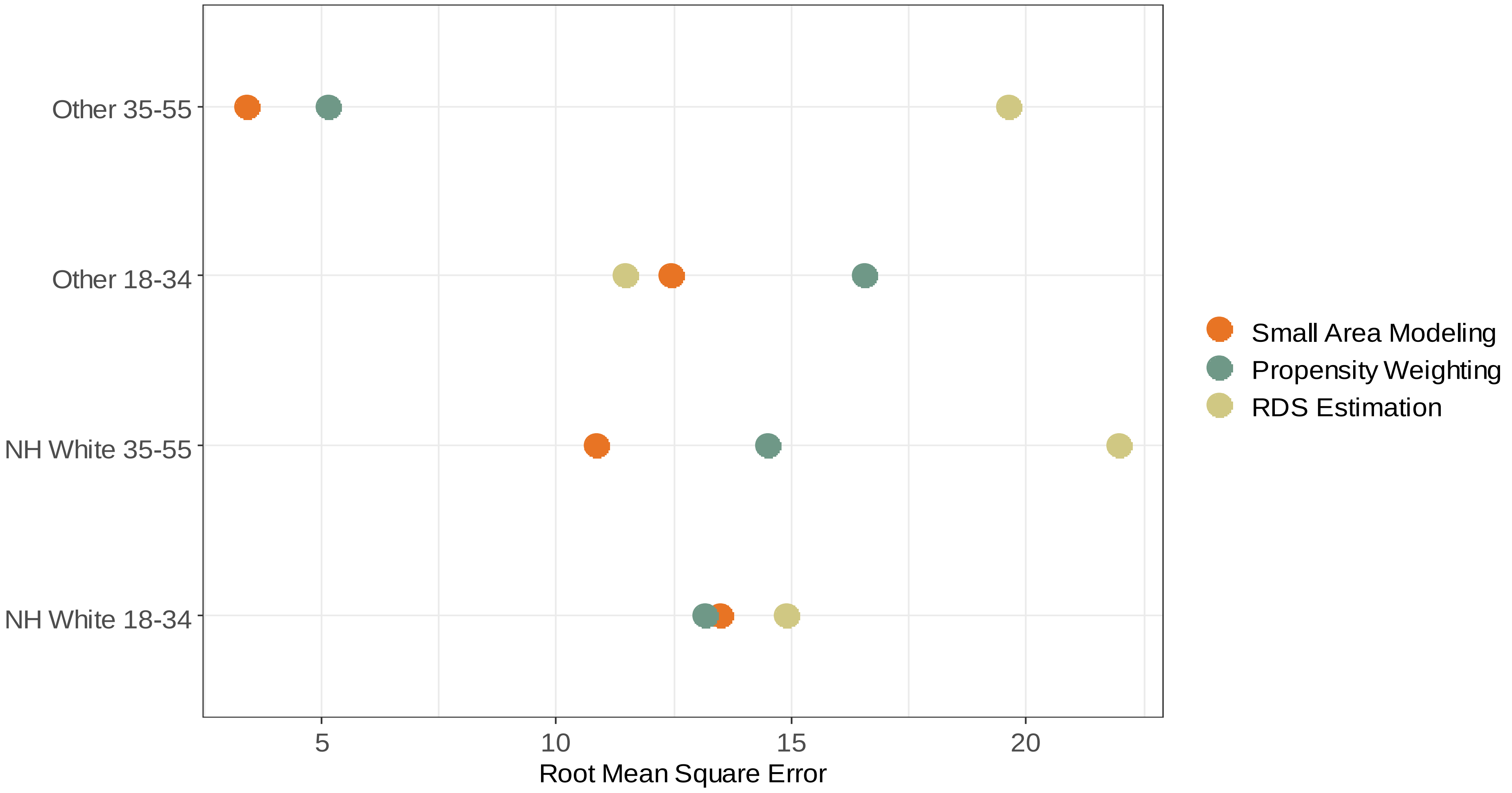
Combined Prob and NonProb Weighted Estimates: Used E-cigarettes in the Last 30 Days

	Estimate	SE	RMSE
RDS Estimation	20.7	2.79	8.15
Small Area Modeling	22.3	3.51	6.97
Propensity Weighting	24.2	4.87	6.38
NHIS 2017	28.4	3.68	

Combined Prob and NonProb Weighted Estimates: Used E-cigarettes in the Last 30 Days by Domain

	RDS Estimation			Small Area Modeling			Propensity Weighting			NHIS 2017 Benchmark	
Domain	**Estimate**	**SE**	**RMSE**	**Estimate**	**SE**	**RMSE**	**Estimate**	**SE**	**RMSE**	**Estimate**	**SE**
NH White 18-34	18.7	4.3	14.88	19.8	3.10	13.50	24.7	10.21	13.17	33.0	5.3
NH White 35-55	9.2	3.8	21.97	20.5	3.19	10.84	18.4	7.46	14.50	30.8	7.7
Other 18-34	31.9	6.0	11.45	33.1	5.91	12.44	36.4	8.38	16.55	22.1	8.5
Other 35-55	24.9	10.6	19.61	10.5	2.66	3.42	8.2	5.15	5.16	8.4	8.4

Combined Prob and NonProb Weighted Estimates: Used E-cigarettes in the Last 30 Days by Race x Age Domains



NORC at the UNIVERSITY of CHICAGO

Observations for Study 3 Smoking Behavior Survey

- Compared to benchmark (NHIS 2017) overall
  - RDS estimation on average had largest MSEs for 3 of 4 domains
  - Propensity weighting had smallest overall MSEs despite largest SEs
  - Small area modeling had smallest average MSEs by domains
- Limitations
  - Small sample size
  - AmeriSpeak sample limited to those with Internet preference
  - Choice of benchmarks
    - **Estimates for LGBT population vary largely from survey to survey**
  - Short questionnaire
    - **Only tested 3 survey variables**

Working Framework for Measuring & Reporting TSE

- **Target Population and Coverage**
  - Seek, use, and report benchmarks to the extent possible for both total and subdomain levels
  - Evaluate and report whether the nonprobability sample covers the target population (e.g. non-Web)
- **Sampling**
  - Probability sample
    - Establish a minimum sample size based on expected sub-domain estimation and analysis
  - Nonprobability
    - Hate them or love them – *Use and report sample quotas*
- **Nonresponse**
  - Probability sample
    - Adjust for potential nonresponse bias using traditional methods
  - Nonprobability
    - Information likely not available for nonresponse bias adjustments, subsumed in estimation approach
- **Estimation for Combined Probability and Nonprobability Sample**
  - Use an approach for inference that adjusts for bias using control totals. E.g. True North Small Area
  - Covariates used in estimation need to be selected carefully and reported
  - Combine bias estimate with standard variance estimate and report MSE for key outcomes.

Future Research

NORC at the UNIVERSITY of CHICAGO

Future Research

- All models rely on covariates and these covariates need to be selected carefully
  - Should modern variable selection mechanisms be used?
  - There are restrictions for methods that rely on the presence of population benchmarks
- Bias Assessment
  - In large nonprobability samples bias is likely to be the most important source of error.
  - The Small Area method uses a probability sample to estimate bias. Can this be done with the other methods?
  - If no companion probability sample is available, use files such as the American Community Survey?
- Mean Squared Error Estimation
  - Use of Total Survey Error techniques to classify and aggregate errors remains the gold standard for transparency and confidence in the data.
    - Need to develop a user-friendly "report card" for clients.
- Composite estimators can be used to combine probability and nonprobability samples

$$\hat{X}_{comb} = \lambda^* \hat{X}_P + (1 - \lambda^*)\hat{X}_{NP}\text{ , where } \hat{X}_P = \sum w_{Pi} X_{Pi} \text{ and } \hat{X}_{NP} = \sum w_{NPi} X_{NPi}$$

NORC at the UNIVERSITY of CHICAGO

Acknowledgement

The authors acknowledge the analytical support provided by the following colleagues:

- Julia Batishev
- Adrijo Chakraborty
- Ying Li
- Lin Liu
- Qiao Ma
- Xian Tao
- Kanru Xiu

Vicki Pineau
pineau-vicki@norc.org

Thank You!

NORC at the UNIVERSITY of CHICAGO

insight for informed decisions™

Selected References

- Bethlehem J. (2015) “Solving the nonresponse problem with sample matching?” *Social Science Computer Review*, Vol. 34, Issue 1, pp. 59 – 77.
- D'Orazio M., Di Zio M., Scanu M. (2006) *Statistical matching: Theory and practice*. Wiley, Chichester.
- Elliot, M. R., Valliant, R. (2017). “Inference for Nonprobability Samples,” *Statistical Science* 2017, Vol. 32, No. 2, 249–264.
- Fahimi, M., Barlas, F.M., Thomas, R.K. and Buttermore, N. (2015). “Scientific surveys based on incomplete sampling frames and high rates of nonresponse,” *Survey Practice*, v.8 (5).
- Fay, R.E., and Herriot, R.A. (1979). “Estimates of income for small places: An application of James-Stein procedures to Census data,” *Journal of the American Statistical Association*, v. 74 (366), pp. 269-277.
- Ganesh, N., Pineau, V., Chakraborty, A., Dennis, J.M., (2017). “Combining Probability and Non-Probability Samples Using Small Area Estimation.” *Joint Statistical Meetings 2017 Proceedings.*
- Rao, J.N.K. (2003). *Small Area Estimation*, John Wiley & Sons, Inc.
- Valliant, R., Dorfman, A. H. and Royall, R. M. (2000). *Finite Population Sampling and Inference*: *A Prediction Approach*. Wiley, New York.
- Ybarra, L.M.R. and Lohr, S.L. (2008). “Small area estimation when auxiliary information is measured with error,” *Biometrika*, v. 95 (4), pp. 919-931.
- Yang, Y. Michael, Nada Ganesh, Ed Mulrow, and Vicki Pineau. (2018). “Estimation Methods for Nonprobability Samples with a Companion Probability Sample,” Proceedings of the Joint Statistical Meetings, 2018

Superpopulation Modeling

- Suppose that the mean of a variable y_i follows a linear model:

  $$E_M(y_i\ /X_i) = X_i'\boldsymbol{\beta}$$

  where X_i is a vector of p covariates for unit i and $\boldsymbol{\beta}$ is a parameter vector.

- Given a sample s, an estimator of the parameter vector is:

  $$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}_s'\boldsymbol{X}_s)^{-1}\boldsymbol{X}_s'y_s$$

- A predictor of the y population total is: $\hat{t} = \sum_{i \in s} y_i + (\boldsymbol{t}_{Ux} - \boldsymbol{t}_{sx})'\widehat{\boldsymbol{\beta}}$ where $\boldsymbol{t}_{Ux}$ and $\boldsymbol{t}_{sx}$ are vectors of X totals for the population and sample, respectively.

- $\hat{t}$ can be written as the weighted sum of the observed y's where the weights are:

  $$w_i = 1 + (\boldsymbol{t}_{Ux} - \boldsymbol{t}_{sx})'(\boldsymbol{X}_s'\boldsymbol{X}_s)^{-1}X_i$$