Encipher Hybrid:

A Validated Methodology for Blending Probability and Nonprobability Samples

This paper introduces the new SSRS Encipher Hybrid methodology for blending probability and nonprobability samples. Encipher is a unique and sophisticated method that leverages study-specific outcomes, advanced modeling techniques, and customized non-demographic measures to produce weighting margins that are optimized for reducing selection bias in key study outcomes. In a validation study whose results are reported here, Encipher:

- Reduced bias in topline estimates by nearly 60% relative to a nonprobability-only sample.
- Reduced bias in subgroup estimates (including breakouts by gender, age, educational attainment, and race) by similar amounts.
- Substantially increased effective sample sizes relative to a probability-only sample.

Hybrid Samples: The "What" and the "Why"

Researchers are increasingly turning to online opt-in sample providers as a lower cost means of collecting survey data. Samples obtained from opt-in sources are **nonprobability samples** because respondents "volunteer" to participate and do not have known selection probabilities. This contrasts with **probability samples**, in which only randomly selected units from a well-defined list of population members are invited to participate.

Because they are recruited from a voluntary sampling pool and surveyed over the Web, **nonprobability samples typically have a lower costs-per-complete than comparable probability samples**. In an era of ever-increasing demand for data-driven decision-making, these cost savings make nonprobability samples enticing for many applications.

However, **whenever decision-making requires insights to be generalized from a specific sample to some larger population, relying solely on nonprobability samples carries significant risks**.

With probability samples, random selection ensures that the selected sample is representative of the population from which it is drawn. Despite recent declines in response rates, probability samples continue to produce accurate population estimates after weighting on demographics.[1]
In contrast, **a nonprobability sample, on its own, is not statistically representative of any target population**. Therefore, **population-level insights may be significantly biased in ways that cannot be corrected with traditional demographic weighting**. Online-only samples exclude individuals who do not use the Internet, but this is only part of the story. Differences between the general population and opt-in

[1] *Keeter et al. 2017; McInnis et al. 2018.*

respondents extend far beyond Web use and observable demographics to encompass attributes like media consumption, shopping habits, life satisfaction, altruism, and civic and political engagement.[2] Consequently, research consistently finds that many nonprobability-based estimates are less accurate than probability-based estimates, even after demographic weighting.[3] These biases are present in general-population estimates and (sometimes to an even greater extent) within key subpopulations.[4]

Enter the SSRS Encipher Hybrid — an ideal solution for researchers who need a "middle ground" between the greater accuracy of probability samples and the lower cost of nonprobability samples. In a hybrid design, we administer a survey to side-by-side probability and nonprobability samples, and then blend the two sets of completes. **The probability sample acts as an "anchor" to allow generalizability to the population, while the nonprobability sample provides a cost-effective source of additional completes**, allowing a larger total sample than could feasibly be obtained from probability sources alone. We apply the SSRS **specialized calibration methodology that matches the nonprobability completes to the probability completes on non-demographic characteristics that are related to key study outcomes**. This corrects for known selection biases and allows the hybrid sample, as a whole, to provide a reasonable snapshot of the target population.

SSRS Encipher Hybrid Methodology: How It Works

Encipher Hybrid offers solutions *throughout the survey lifecycle* — before, during, and after data collection — to allow probability and nonprobability completes to be analyzed as a single sample that can accurately be generalized to the population of interest.

Before Data Collection

We select a handful of **topic-customized items from the SSRS Encipher Calibration Item Bank** for inclusion on the questionnaire. Including these items allows us to go beyond simple demographic weighting and one-size-fits-all solutions to develop a weighting model that is well-tailored to study-specific measures.

Our Calibration Item Bank includes about 40 (and growing) non-demographic items that have been experimentally validated as being strong predictors of differences between probability and nonprobability samples. The items cover multiple topic areas, including Internet and Technology Use, Consumer Behavior, Political Attitudes, Health Behavior, Social and Institutional Trust, Privacy Attitudes, Science Attitudes and Knowledge, and Sports and Leisure Activities. This allows us to select items that are *customized to the topic of a given study*. Topic customization is critical because, to meaningfully reduce the risk of bias, weighting variables must be predictive of the substantive measures for which population estimates are desired.[5]

---

[2] *DiSogra et al. 2011; Fahimi et al. 2015.*
[3] *Yeager et al. 2011; MacInnis et al. 2018; Cornesse et al. 2020; Pasek and Krosnick 2020.*
[4] *Kennedy et al. 2016.*
[5] *Little and Vartivarian 2005.*

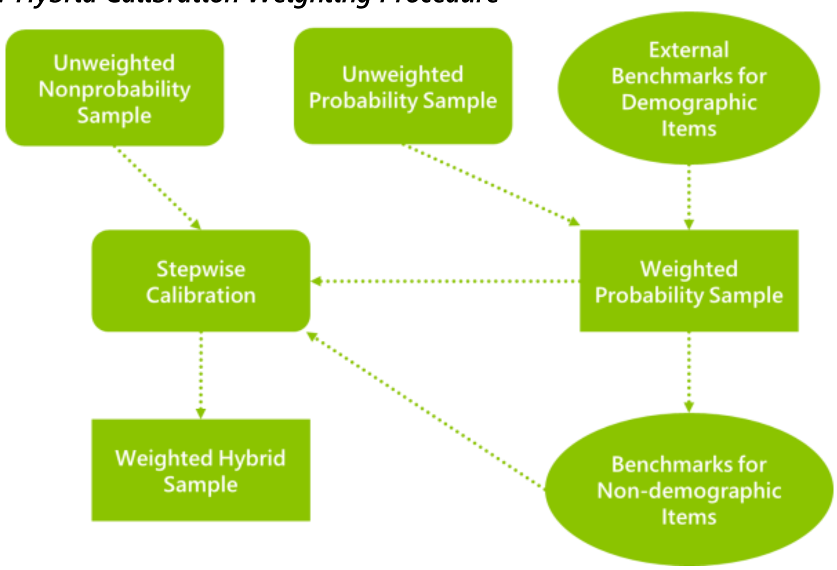
During Data Collection

We administer the full questionnaire to **side-by-side probability and nonprobability samples**. The size of the probability sample is customized to study needs, but typically comprises 25% to 50% of the total completes. For many target populations, our SSRS Opinion Panel is available as a cost-effective source of probability completes. Of course, SSRS also offers custom probability-based designs.

After Data Collection

After data collection is complete, figure 2 illustrates the process we follow, beginning with the unweighted probability and nonprobability samples, to produce a calibrated hybrid weight.

Figure 2: Encipher Hybrid Calibration Weighting Procedure



We begin by weighting the probability sample to an extended set of external demographic benchmarks. We use this weighted probability sample to produce internal benchmarks for the non-demographic items from the Calibration Item Bank.

We then apply the SSRS **Stepwise Calibration methodology to develop calibrated hybrid weights** adjusted both to the external demographic benchmarks and the most useful internal calibration benchmarks. Stepwise Calibration adapts a guided-search algorithm[6] to identify a final set of weighting margins that is optimal for minimizing bias across a pre-identified set of key substantive measures from the survey. The algorithm begins with a large set of potential weighting margins that includes the items from the Calibration Item Bank along with key demographics. It then progressively narrows this "search space" to a much more parsimonious weighting model, retaining only those margins that meaningfully contribute to reducing selection bias. Using Stepwise Calibration, SSRS data scientists can develop study-tailored weighting models that balance bias reduction and other data quality measures across multiple study outcomes, all while keeping budget and turnaround time under control.

[6] *For examples of similar approaches, see Schouten 2007 and Särndal and Lundström 2010.*

Validation of the SSRS Encipher Hybrid Approach

Study Design

To validate the Encipher Hybrid methodology, we fielded several surveys covering a broad range of topic areas. Each survey included (1) relevant items from the Encipher Calibration Item Bank and (2) several benchmarkable outcome items that we used to evaluate the success of calibration at reducing selection bias.

Each survey was fielded to:

- A probability sample selected from the SSRS Opinion Panel. Most completes from this sample were by Web, with some phone completes to represent non-Web users.
- A nonprobability sample purchased from an opt-in Web panel vendor. All completes from this sample were by Web.

We combined the resulting completes to create a hybrid sample in which 25% of the completes were from the probability-based SSRS Opinion Panel and the remaining 75% were from nonprobability sources. We applied the Encipher methodology to calibrate on non-demographic items from the Calibration Item Bank, in addition to standard demographics. We refer to this design as *Hybrid – calibrated*.

We then compared estimates from the *Hybrid – calibrated* design to estimates from three other designs, all of which were weighted only on standard demographics: *Probability*, using only the SSRS Opinion Panel; *Nonprobability – demo*, using only the nonprobability completes; and *Hybrid – demo*, using the same 25%-75% split but omitting Encipher calibration.

Results: Selection Bias

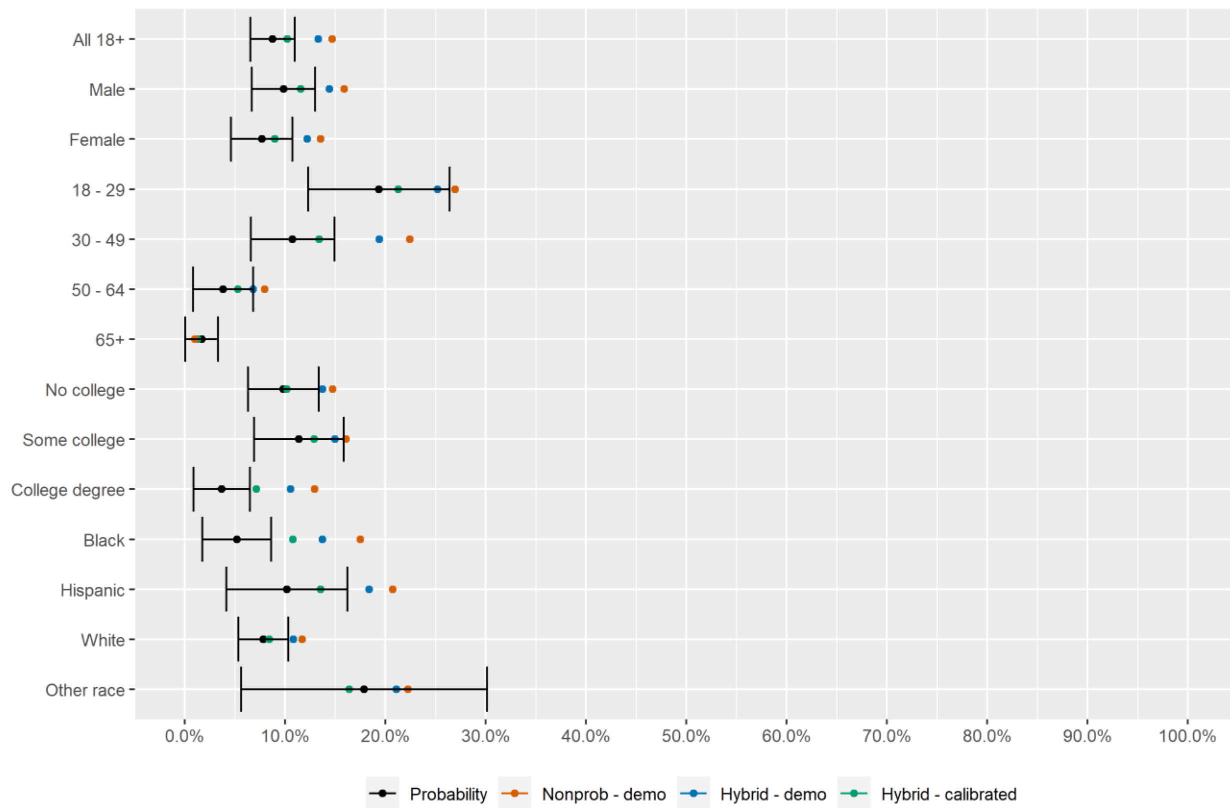
Example Outcome: E-cigarette Use

Figure 3 shows one of the outcome estimates—the percent of adults who use e-cigarettes or other vaping products—under each of these four designs. The figure includes the topline estimate for the entire 18+ U.S. population as well as breakouts by gender, age, education, and race/ethnicity categories.

The hybrid sample using SSRS's Encipher methodology (*Hybrid–calibrated*) achieves topline and subgroup estimates that are, in most cases, indistinguishable from the *Probability* estimates. Specifically, for the topline and most subgroup estimates, the *Hybrid–calibrated* estimate is within the 95% confidence bounds of the *Probability* estimate.

In contrast, relying on a nonprobability sample alone (*Nonprobability – demo*) would cause us to significantly overestimate the percent of adults using e-cigarettes, relative to the Probability benchmark. A hybrid that weights only on demographics, without Encipher calibration (*Hybrid – demo*), would be closer to the Probability benchmark than the nonprobability-only estimate; but in most cases, the estimate would remain outside the 95% confidence bounds, and in some cases, well outside.

Figure 3: Percent Using E-cigarettes, by Sampling/Weighting Methodology



NOTE: Error bars show the 95% confidence interval around the Probability estimate.

All Outcomes

Figure 4 generalizes these results, plotting the observed selection bias for all outcomes collected in the validation study. Again, results are shown both for topline (all 18+ adults) estimates and for breakouts by common demographic categories.

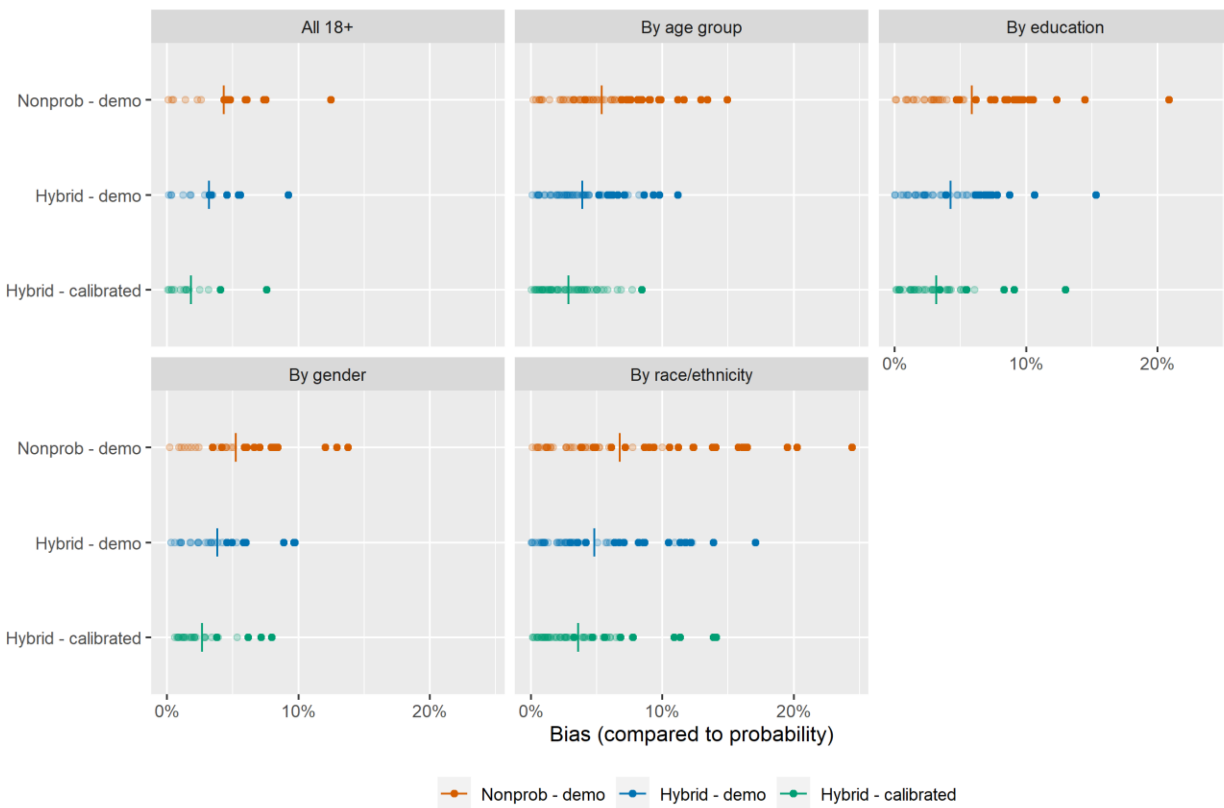
In relative terms, for topline estimates, the **SSRS Encipher Hybrid methodology reduces the average selection bias by nearly 60% compared to nonprobability-only designs**, and by **over 40% compared to hybrid designs that weight only on demographics**. For many outcomes, after applying Encipher, the hybrid estimate is within the 95% confidence bounds of the corresponding Probability estimate.

Encipher also reduces the variability in the observed bias across outcomes. This is important because research shows not only that nonprobability samples are more biased on average than probability samples, but also that they show greater variability in bias, leading to greater uncertainty as to the likely accuracy of any specific estimate.[7] Conversely, Encipher Hybrid, by reducing this variability, can make users more confident that a given estimate is representative of the target population.

[7] *Dutwin and Buskirk 2017.*

Finally, **the consistent advantage of Encipher Hybrid holds both for topline estimates and for estimates broken out by demographic subgroups**, including subgroups that are commonly of high interest to researchers (e.g., Hispanic adults, younger age groups, and adults without any college education). As a companion to figure 4, table A.1 in the Appendix shows the average bias, maximum bias, and number of outcomes with statistically significant bias under each methodology for specific subgroups.

Figure 4: Selection Bias in Validation Study Outcomes, by Sampling/Weighting Methodology



NOTE: Vertical line shows the average bias across outcomes. Shading indicates statistical significance: light-shaded estimates are within the 95% confidence bounds of the corresponding Probability estimate, while dark-shaded estimates are outside the 95% confidence bounds.

Results: Effective Sample Sizes

Though complex sample designs and weighting procedures can help reduce selection bias, they can also increase the variability of weights, leading to larger margins of sampling error.[8] The impact of the sample design and weighting on sampling error can be represented by the **effective sample size**. This is the sample

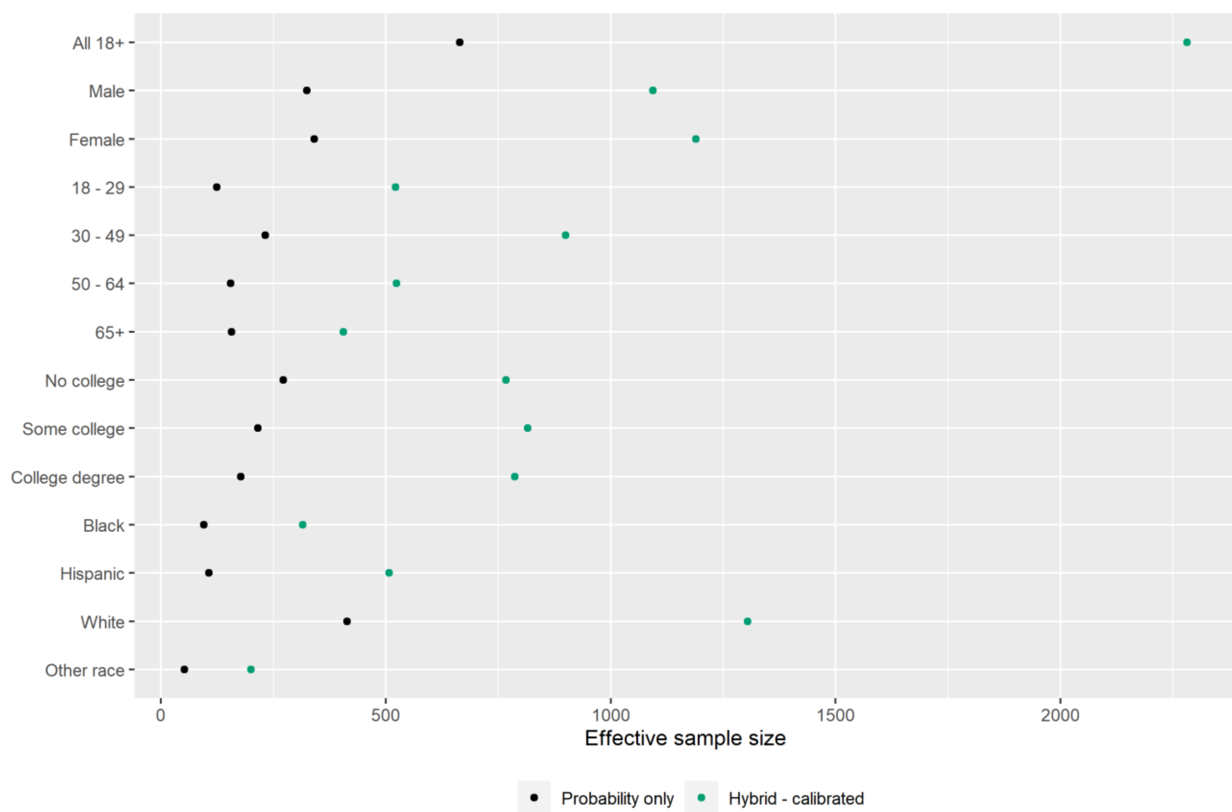
[8] *The margin of sampling error, as commonly reported, is simply half the width of a 95% confidence interval. The larger the margin of sampling error, the more the estimate would be expected to vary across samples if the study were repeated many times with the same sampling and weighting procedures. Generally, the margin of sampling error increases as the sample size decreases and/or as the weights become more variable.*

size on which the margin of error is computed and accounts for the reduction in precision resulting from the complex design and weighting.

The benefit of any hybrid design (relative to a probability-only design) is that the additional nonprobability completes increase the total sample size, which (all else equal) would lead to smaller margins of sampling error. However, inefficient or ineffective weighting can offset this benefit of the larger sample. If this were the case, it would be reflected in a lower effective sample size.

To assess whether this is the case with Encipher, figure 5 shows effective sample sizes, overall and within subgroups, from the Health Behavior module of the validation study for the *Probability* and *Hybrid–calibrated* designs. This illustrates that **the hybrid design substantially increases effective sample sizes overall and within all key subgroups**. Thus, **the Encipher calibration effectively reduces bias without offsetting the benefit of the additional nonprobability sample**.

Figure 5: Effective Sample Size, Hybrid–calibrated vs. Probability Only



Conclusion

As demonstrated by this validation study, SSRS's Encipher – Hybrid methodology allows researchers to increase effective sample sizes while reducing the risk of serious selection bias. Instead of relying solely on nonprobability data, a hybrid design using SSRS's Encipher calibration methodology can increase researchers' confidence that their results accurately represent the population of interest. In this way, hybrid designs can offer an affordable "middle ground" between the greater accuracy of probability samples and the lower cost of nonprobability samples, particularly for harder-to-reach populations for which a purely probability-based sample may be cost prohibitive.

For more information about how Encipher Hybrid could be useful for your study, visit https://ssrs.com/encipher-hybrid/.

Appendix: Bias by Individual Subgroups

Table A.1. Average, maximum, and statistical significance of bias, by sampling/weighting methodology

Subgroup	Average bias			Maximum bias			Number of estimates with statistically significant bias[1]		
	Nonprobability–demo	Hybrid–demo	Hybrid–calibrated	Nonprobability–demo	Hybrid–demo	Hybrid–calibrated	Nonprobability–demo	Hybrid–demo	Hybrid–calibrated
All 18+	4.3%	3.2%	1.8%	12.5%	9.2%	7.6%	9	7	2
Male	4.8%	3.6%	2.5%	12.9%	9.7%	8.0%	5	3	1
Female	5.6%	4.1%	2.9%	13.8%	9.7%	7.2%	9	7	4
18 - 29	6.0%	4.4%	3.3%	11.2%	8.2%	6.6%	5	2	0
30 - 49	5.9%	4.3%	2.9%	13.4%	9.8%	8.4%	5	5	1
50 - 64	6.4%	4.6%	3.0%	14.9%	11.2%	7.7%	5	4	0
65+	3.3%	2.3%	2.2%	9.8%	7.2%	6.9%	2	0	0
No college	6.8%	4.9%	3.4%	12.3%	8.8%	8.3%	10	7	2
Some college	4.0%	3.0%	2.5%	14.5%	10.6%	9.1%	4	2	2
College degree	6.8%	4.9%	3.4%	20.9%	15.3%	13.0%	7	5	2
Black	5.2%	3.6%	3.2%	16.4%	11.8%	11.4%	4	2	3
Hispanic	5.3%	3.9%	3.1%	13.8%	10.5%	7.8%	4	3	1
White	5.3%	3.9%	2.7%	16.5%	12.2%	10.9%	9	7	3
Other race	11.4%	7.9%	5.3%	24.5%	17.1%	14.1%	7	3	2

[1] A Nonprobability or Hybrid estimate is considered to show statistically significant bias if it is within the 95% confidence bounds of the corresponding Probability estimate.

References

Cornesse, C., Blom, A.G., Dutwin, D., Krosnick, J.A., De Leeuw, E.D., Legleye, S., Pasek, J., Pennay, D., Phillips, B., Sakshaug, J.W., Struminskaya, B., and Wenz, A. (2020). "A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research." *Journal of Survey Statistics and Methodology 8*(1): 4 – 36.

DiSogra, C., Cobb, C., Chan, E., and Dennis, J.M. (2011). "Calibrating Non-Probability Internet Samples with Probability Samples Using Early Adopter Characteristics." Proceedings of the American Statistical Association Section on Survey Research Methods. Retrieved February 7, 2022 from http://www.asasrms.org/Proceedings/y2011/Files/302704_68925.pdf.

Dutwin, D. and Buskirk, T.D. (2017). "Apples to Oranges or Gala versus Golden Delicious? Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples." *Public Opinion Quarterly 81*(1): 213 – 239.

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*. Retrieved February 7, 2022 from https://www.surveypractice.org/article/2827-scientific-surveys-based-on-incomplete-sampling-frames-and-high-rates-of-nonresponse.

Keeter, S., Hatley, N., Kennedy, C., and Lau, A. (2017). "What Low Response Rates Mean for Telephone Surveys." Pew Research Center. Retrieved February 7, 2022 from https://www.pewresearch.org/methods/2017/05/15/what-low-response-rates-mean-for-telephone-surveys.

Kennedy, C., Mercer, A., Keeter, S., Hatley, N., McGeeney, K., and Gimenez, A. (2016). "Evaluating Online Nonprobability Surveys." Pew Research Center. Retrieved February 7, 2022 from https://www.pewresearch.org/methods/2016/05/02/evaluating-online-nonprobability-surveys.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Little, R. and Vartivarian, S. (2005). "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology 31*(2): 161 – 168.

MacInnis, B., Krosnick, J.A., Ho, A.S., and Cho, M. (2018). "The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension." *Public Opinion Quarterly 82*(4): 707 – 744.

Pasek, J. and Krosnick, J.A. (2020). "Relations Between Variables and Trends Over Time in RDD Telephone and Nonprobability Sample Internet Surveys." *Journal of Survey Statistics and Methodology 8*(1): 37 – 61.

Särndal, C. and Lundström, S. (2010). "Design for Estimation: Identifying Auxiliary Vectors to Reduce Nonresponse Bias." *Survey Methodology 36*(2): 131 – 144.

Schouten, B. (2007). "A Selection Strategy for Weighting Variables Under a Not-Missing-at-Random Assumption." *Journal of Official Statistics 23*(1): 51 – 68.

Yeager, D.S., Krosnick, J.A., Chang, L., Javitz, H.S., Levendusky, M.S., Simpser, A., and Wang, R. (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.