GENESYS

Marketing Systems Group

HYBRID SAMPLING 2.0

Pushing the Survey Sampling Envelope in the Digital Age

Introduction

The survey sampling landscape is rapidly evolving. In an era of diminishing response rates, escalating data collection costs, and shifting communication preferences, more effective survey sampling alternatives are no longer academic curiosities [1]. While the new realities suggest that departures from traditional methods are becoming inevitable, they also beckon an immediate question as survey researchers continue to experiment with hybrid sampling techniques. That is,

> *Are such sampling alternatives conducive to the inferential integrity of scientific surveys by reaching a representative subset of the target population in a pragmatic and cost-effective manner?*

In addition to adopting multiple modes of data collection [2], it has become a customary practice to use less expensive samples selected from online panels to supplement costly alternatives from address or telephone frames [3]. The so-called opt-in panels are compiled using a potpourri of recruitment techniques, mostly relying on social media to "fish" for individuals willing to partake in surveys – hence the term river sampling [4]. The resulting convenience of these recruitment methods, however, is often achieved at the expense of compromising the organic representation that has been a natural byproduct of probability samples. This trade-off becomes of elevated concern since with samples obtained from opt-in panels, typical geodemographic weighting adjustments may no longer be adequate for ensuring their representativity [5].

It has been suggested that with such samples more granular weighting and calibration adjustments become necessary to ameliorate their compromised representations [6]. Specifically, calibration adjustments to behavioral and attitudinal benchmarks that go beyond geodemographic corrections might be needed to improve the representation of survey respondents from less representative samples [7,8]. Moreover, when mixing samples secured from different sampling frames, special procedures must be used to combine the various sample components in an optimal fashion [9]. If conducted effectively, the resulting hybrid samples may address the cost challenge of traditional surveys that rely on a single frame for sampling [10].

However, manufactured representation via heavy weighting and calibration adjustments of samples from opt-in panels may fail to render the level of precision scientific surveys demand. Moreover, sample supplementation using such panels can quickly become infeasible when the survey geography is small, the target population is of low incidence, or when hard-to-get cohorts are of interest. Lastly, these alternative sampling methods are becoming increasingly vulnerable to fraudsters who pose as real survey-takers to collect incentives that have to be offered to improve participation rates.

In light of the above, the main research question this paper aims to address is that: can advantage be taken from the emerging digital technologies to recruit survey respondents with improved representation of their target population as compared to what the existing opt-in panels can provide? In particular, can an innovative hybrid sampling strategy offer pragmatic solutions for small area surveys where other sampling alternatives of convenience fail to be feasible?

Methodology

As mentioned earlier, current hybrid sampling options whereby probability sample components are supplemented with those from opt-in panels are subject to serious quality shortfalls due to their inherent unknowable representation and vulnerability to fraudulent respondents. Our proposed methodology introduced here relies on digital advertisements that are reference-aligned vis-à-vis the socioeconomic characteristics of the survey population to offer supplementation solutions for hybrid samples that have a core probability-based component. This avant-garde supplementation strategy, which is more robust to intrusion by bots and kick-farms, is particularly attractive for small areas where no other sample supplementation alternatives can offer feasible options.

Our proposed methodology can recruit survey respondents directly from the websites and platforms they access, across a vast array of online networks. While most digital advertising campaigns are designed to find similar audiences to increase clicks for sales optimization, irrespective of any concerns for representation, our methodology circumvents greedy algorithms that aim to recruit more of the same. This improvement in diversity and representation, which reduces bias due to snowballing effects, is achieved by actively imposing and managing a myriad of geodemographic and socioeconomic quotas. This means our dynamic sample recruitment management ensures that prespecified quota are filled with respondents who reflect the population of interest.

Specifically, the above quasi-probability sampling approach pursuits respondents using stratified digital advertising through a combination of online and social media platforms. The recruitment process follows a quota-based sample management specific to the design needs of studies [11]. Quotas are monitored and adjusted dynamically throughout the field period to maintain alignment with population benchmarks while undergoing real-time quality assurance and fraud detection procedures.

Case Study

In 2024, the Tarrant County Public Health and the Centers for Disease Control and Prevention conducted a telephone survey to update the various health metrics of the county residents. This survey, which was designed to follow the protocols of the Behavioral Risk Factor Surveillance System (BRFSS) [12] relied on the method of Dual Frame Random Digit Dialing (DFRDD) for data collection. As such, a total of 3,512 adults residing in Tarrant County of Texas were surveyed. Subsequently, respondents of this survey were weighted to represent the target population, by making adjustments that aimed to correct coverage issues due to sampling and differential nonresponse.

In order to contrast the performance of our proposed sampling method against the best conventional alternatives can offer, the above survey was replicated using our hybrid sampling protocol. This protocol was comprised of a core probability sample of 477 respondents secured from an Address-Based Sampling (ABS) survey of the County, supplemented with 488 respondents recruited via our digital recruitment platform (MSGD). Moreover, the same ABS core component was also supplemented with 500 respondents from opt-in panels to serve as status quo alternative (MSGO). That is, two sets of hybrid sample surveys were conducted to compete against the DFRDD survey conducted by Tarrant County. These two sets were:

1. ABS + MSGD, with 965 = 477 + 488 combined respondents.
2. ABS + MSGO, with 977 = 477 + 500 combined respondents.

The MSGD sample component was recruited to boost the type of respondents that traditional ABS push-to-web was anticipated to experience coverage challenges. This included households for which commercial databases have limited information about. The MSGO component included all the respondents panel companies could recruit via the Exchange, without any input from MSG.

It should be noted that both the ABS and each of the above two supplemental sample surveys used items (questions) borrowed from the Tarrant County BRFSS questionnaire. In addition to a common battery of demographic questions embedded to support weighting adjustments, each survey included a series of behavioral and attitudinal questions to support comparative analyses. The latter questions included:

- Are you on active duty?
- How often do you wear seatbelts?
- Do you have difficulty climbing stairs?
- Have you smoked at least 100 cigarettes so far?
- How is your general health?
- How is your general mental health?
- How affordable is medical support?

- When was your last medical check-up?
- When was your last dental check-up?
- Do you have depression?
- Do you have asthma?

Analysis

Comprehensive evaluation of the efficacy of any alternative method, sampling or otherwise, has to be measured vis-à-vis existing options that are currently in use. Furthermore, such evaluations have to rely on multivariate metrics that consider not only the product quality but also the cost and timeliness of the results under the proposed alternative methodology. While this perspective does not imply that significant cost and time savings can trump or justify poor data quality, comparisons that do not reflect such ground realities would be merely academic and void of pragmatic relevance.

In order to adhere to the above, our comparative analyses set out to assess the quality of the resulting survey estimates from each of the above two hybrid sample surveys against what the conventional County survey had produced. Furthermore, cost and time comparisons were carried out relative to what a typical DFRDD of comparable scope would require. While the latter two comparisons were rather straightforward, the process used for comparing the quality of the resulting surveys included the following steps:

Weighting Adjustments – All survey data must be weighted before they could be used to produce unbiassed estimates of population parameters. By improving the representation of respondents, weighting reduces bias and enhances the external validity of survey estimates. The weighting process for each sample survey followed what was used for the DFRDD survey. That is:

1. In the first step, design weights were computed to reflect selection probabilities that included surveying only one adult per household.
2. In the second step, design weights were calibrated to the demographic distributions of the target population for whom the needed benchmarks were obtained from the latest American Community Survey (ACS 2023). These calibration adjustments were carried out using the WGTADJUST procedure of SUDAAN to balance the distributions of survey respondents against multiple benchmarks simultaneously. This procedure relies on a constrained logistic regression to predict the likelihood of response, which then will be used to create adjustments that align respondents to the specified benchmark distributions [13]. The ACS benchmarks used for weighting adjustments were doubly indexed by gender and the following demographic characteristics:

   a. Age: 18 - 24, 25 - 34, 35 - 44, 45 - 54, 55 - 64, 65 - 74, and 75+
   b. Ethnicity: Hispanic and Other

c. Race: White, Black, Asian, and Other
d. Education: No HS, HS, Some College, AS, BS, and MS+
e. Income: $0K<$25K, $25<$50K, $50K<$75K, $75K<$100K, $100K<$150K, $150K<$200K, and $200K+
f. Marital: Married, Widowed, Divorced, Separated, and Never Married
g. Adults: 1, 2, 3, 4, and 5+

3. In the third step, produced weights were examined to identify and ameliorate extreme values. Trimming extreme weights is a standard practice that is used to improve the efficiency of the weighting process and add stability to survey estimates. To maintain consistency, similar rules were used to trim large and small weights for each survey.

Quality Assessment – Once respondents from the core ABS and their corresponding hybrid samples were weighted using an identical calibration process outlined above, the mean square error ratio (MSER) for each of attitudinal/behavioral items were calculated separately for each sample survey. Arguably, MSER is a more telling indicator since it considers both the negative effect of variance inflation due to weighting as well as its positive gain for bias reduction [14]. For this purpose, estimates obtained from the DFRDD survey were considered of the highest quality and served as reference points. As such, for each estimate from the ABS and competing hybrid samples a measure of dispersion was computed by:

$$MSER(\hat{p}_{competing}) = \frac{MSE(\hat{p}_{competing})}{MSE(\hat{p}_{DFRDD})}$$

Expanding the above would result in:

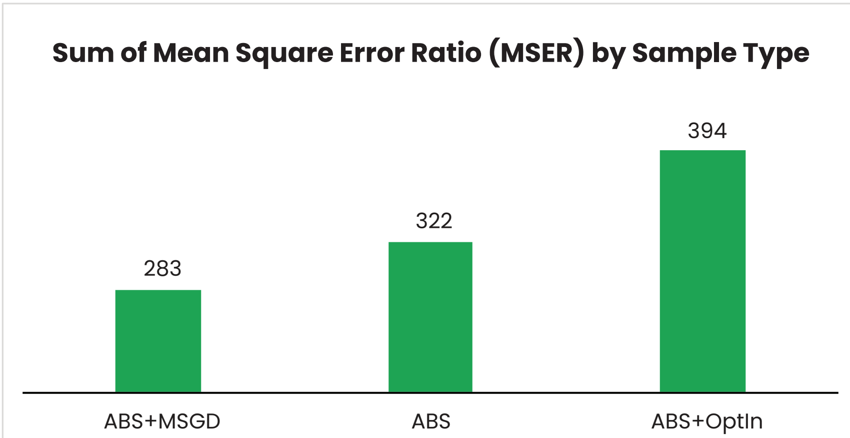
$$MSER(\hat{p}_{competing}) = \frac{Bias^2(\hat{p}_{competing}) + Var(\hat{p}_{competing})}{Bias^2(\hat{p}_{DFRDD}) + Var(\hat{p}_{DFRDD})}$$

Assuming that DFRDD estimates are unbiased, the above would simplify to:

$$MSER(\hat{p}_{competing}) = \frac{Bias^2(\hat{p}_{competing}) + Var(\hat{p}_{competing})}{0 + Var(\hat{p}_{DFRDD})}$$

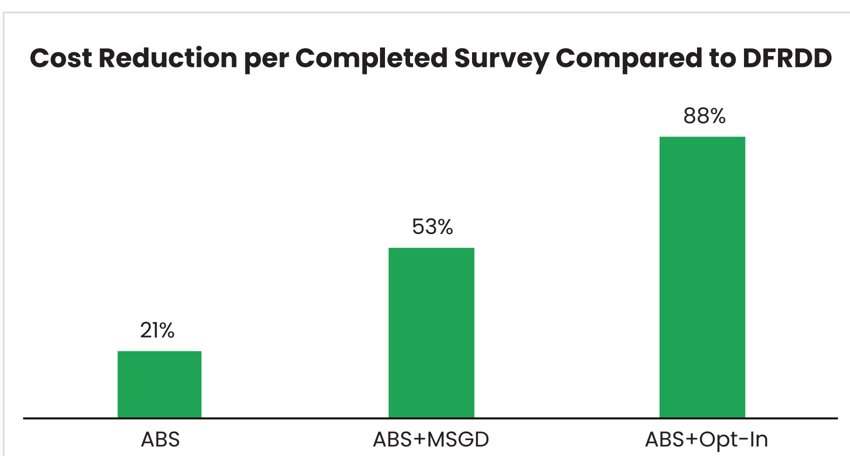
Results

With an MSER computed for each comparison item of each survey, the sum of these statistics were computed to represent the overall deterioration metric of quality of their resulting survey estimate in reference to what the DFRDD has been able to produce. The following chart provides a summary of these metrics for the two hybrid samples, as well as that for ABS sample alone.



Clearly, a hybrid sample that is comprised of an ABS core supplemented with respondents recruited digitally using our proposed method (MSGD) tend to produce survey estimates that are superior to what supplementation via opt-in panels can offer. In part, this reflects one of the key limitations of opt-in panels in smaller geographies where very little quota balancing can be enforced upfront to improve the representation of its respondents.

As mentioned at the outset, any evaluation of alternative sampling methods will be incomplete if their cost ramifications are not considered. Such assessments will be equally incomplete if coverage issues of traditional methods resulting from their punishing rates of nonresponse and frame issues are overlooked. To illustrate, the following chart shows the reduction in cost for completing one survey under the ABS and the two sample supplementation methods relative to that of DFRDD.



Conclusions

The new realities of the digital age have brought about a number of resilient challenges for survey researchers, which in turn have exposed some of the inefficiencies associated with the traditional methods this community has relied upon for decades. The resilience of such challenges suggests that piecemeal approaches that may have limited accessibility or scalability will prove to be inadequate and transient. It is from this perspective that our proposed method of sampling has aimed to introduce a durable and accessible solution for hybrid sample surveys.

Increasingly, survey researchers are compelled to rely on hybrid samples to improve coverage, reduce cost, and increase the number of respondents needed for producing robust estimates. As such, researchers often supplement probability samples with those from online panels that are significantly less costly. These panels, however, are mostly developed for commercial applications as they are void of organic representation of any geography. This shortfall becomes particularly acute in smaller areas where opt-in panels have inadequate coverage, or where quota balancing cannot be implemented to synthesize representation – albeit coarsely. Traditional methods of survey sampling are becoming progressively inefficient, both with respect to cost and data quality, begging for novel and pragmatic alternatives that do not forego rigor. Our proposed method of digital recruitment, which is target population aligned, offers a cost-effective sample supplementation alternative that can improve the quality of survey estimates from hybrid samples as compared to frugal options that are currently available to researchers. Results from the detailed comparisons we have exhibited in this paper show that our proposed sampling methodology can produce estimates that are superior in quality than what the current options can provide.

As parting notes, we would like to submit the following considerations as survey researchers continue to push the envelope in pursuit of more effective sampling methods:

a. Despite the growing challenges facing the survey research community, practitioners should not succumb to suboptimal practices for cost-saving purposes alone. Such unilateral guidelines have contributed to the stigma that commercial surveys have inadequate concern for rigor.

b. On the other hand, undue allegiance to traditional methods of survey sampling can also confine researchers to inefficient practices that are losing their pragmatism. This position becomes particularly untenable when such adherences are simply for the sake of preserving the optics.

c. With declining response rates, surveys require progressively more comprehensive weighting adjustments to restore the representation of their respondents. As such, it is not advisable to shy away from more aggressive weighting and calibration adjustments only to keep the resulting variance inflation due to weighting at bay [13]. Of course, this is fully cognizant of the proverbial seesaw occupied on one side by bias and variance on the other.

d. Above all, it is imperative to retain full transparency about adopted methodologies and their potential shortfalls as we explore new possibilities for survey sampling in the digital age.

References

1. Brick JM. The future of survey sampling. Public Opinion Quarterly. 2011;75(5):872-888. *[FREE Full text]* *[doi: 10.1093/poq/nfr045]*
2. Brick JM, Lepkowski JM. Multiple mode and frame telephone surveys. In: Lepkowski JM, Tucker NC, Brick JM, de Leeuw ED, Japec L, Lavrakas PJ, et al, editors. Advances in Telephone Survey Methodology. New York. Wiley; 2008;149-169.
3. Callegaro M, Manfreda KL, Vehovar V. Web Survey Methodology. Los Angeles. Sage Publications; 2015.
4. Baker R, Brick JM, Bates NA, Battaglia M, Couper MP, Dever JA, et al. Report of the AAPOR task force on non-probability sampling. J Surv Stat Methodol. Jun 2013;1(2):1-125. *[FREE Full text]*
5. Mercer A, Lau A, Kennedy C. For weighting online opt-in samples, what matters most? Pew Research Center. 2018. URL: https://www.pewresearch.org/methods/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/ [accessed 2024-01-24]
6. Fahimi M, Barlas FM, Thomas RK, Buttermore N. Scientific surveys based on incomplete sampling frames and high rates of nonresponse. Surv Pract. 2015;8(6):1-11. *[FREE Full text]* [doi: 10.29115/sp-2015-0031]
7. Kott PS. Using calibration weighting to adjust for nonresponse and coverage errors. Surv Methodol. 2006;32(2):133-142. *[FREE Full text]*
8. Lee S. Propensity score adjustment as a weighting scheme for volunteer panel web surveys. J Off Stat. 2006;22(2):329-349. *[FREE Full text]*
9. Fahimi M. Optimal integration of surveys. The University of New Mexico. 1989. URL: https://www.unm.edu/ [accessed 2024-01-24]
10. Fahimi, M., et al. Improving the Efficiency of Inferences from Hybrid Samples for Effective Health Surveillance Surveys: Comprehensive Review of Quantitative Methods. JMIR Public Health Surveill, 10: e48186 (2024).
11. EveryAnswer® is supported by RA2 Inc. and MSG. The platform has been developed through their collaboration, and both organizations contribute expertise to its ongoing growth and operation.
12. Behavioral Risk Factor Surveillance System (BRFSS). Centers for Disease Control and Prevention. URL: http://www. cdc.gov/brfss/index.html [accessed 2024-01-24] RTI International (2012).
13. SUDAAN Language Manual, Release 11.0. RTI International. www.rti.org/sudaan
14. Levey PS, Lemeshow S. Sampling of Populations: Methods and Applications. New York. Wiley; 1991.
15. Skinner C. Calibration weighting and non-sampling errors. Res Official Stat. 1999; 2:33-43.