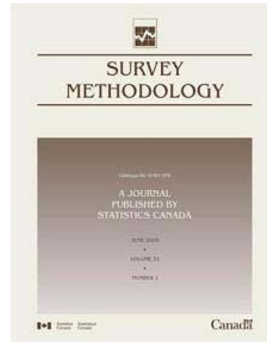
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Use of a web-based convenience sample to supplement a probability sample



by Marc N. Elliott and Amelia Haviland

December, 2007

Statistics Canada Statistique Canada

Canada

Survey Methodology, December 2007
Vol. 33, No. 2, pp. 211-215
Statistics Canada, Catalogue no. 12-001-X

Use of a web-based convenience sample to supplement a probability sample

Marc N. Elliott and Amelia Haviland [1]

Abstract

In this paper we describe a methodology for combining a convenience sample with a probability sample in order to produce an estimator with a smaller mean squared error (MSE) than estimators based on only the probability sample. We then explore the properties of the resulting composite estimator, a linear combination of the convenience and probability sample estimators with weights that are a function of bias. We discuss the estimator's properties in the context of web-based convenience sampling. Our analysis demonstrates that the use of a convenience sample to supplement a probability sample for improvements in the MSE of estimation may be practical only under limited circumstances. First, the remaining bias of the estimator based on the convenience sample must be quite small, equivalent to no more than 0.1 of the outcome's population standard deviation. For a dichotomous outcome, this implies a bias of no more than five percentage points at 50 percent prevalence and no more than three percentage points at 10 percent prevalence. Second, the probability sample should contain at least 1,000-10,000 observations for adequate estimation of the bias of the convenience sample estimator. Third, it must be inexpensive and feasible to collect at least thousands (and probably tens of thousands) of web-based convenience observations. The conclusions about the limited usefulness of convenience samples with estimator bias of more than 0.1 standard deviations also apply to direct use of estimators based on that sample.

Key Words: Bias; Composite estimator; Calibration.

1. Introduction

Web-based surveys have steadily increased in use and take a variety of forms (Couper 2000). For instance, web-based probability samples use a traditional sampling frame and provide web-mode as one response option or the only response option. Web-based probability samples can have high response rates and produce estimators with minimal non-response bias (Kypri, Stephenson and Langley 2004). In contrast, web-based convenience samples are based on "inbound" hits to web pages obtained from anyone online who finds the site and chooses to participate (sometimes as a result of advertising to a population that is not specifiable) or based on volunteerism from recruited panels that are not necessarily representative of the intended population.

The primary appeal of web-based convenience samples lies in the potentially very low marginal cost per case. Visits to a web site do not require expensive labor (as for phone calls) or materials (as for mailings) for each case, combined with rapid data collection and reductions in marginal data processing costs per case. Even with some fixed costs, the total costs per case are potentially very low, especially for large surveys. The disadvantage of these samples is also clear: potentially large and unmeasured selection bias.

Most discussions of web-based convenience samples of which we are aware have either argued that probability samples are unimportant in general, tried to delineate the circumstances under which convenience samples may be useful, or dismissed the use of convenience samples entirely. We explore a different avenue by investigating the possibility of integrating web-based convenience samples into the context of probability sampling.

In this paper we describe a methodology for combining a convenience sample with a probability sample to produce an estimator with a smaller mean squared error (MSE) than estimators that employ only the probability sample. We then explore the properties of the resulting composite estimator, a linear combination of the convenience and probability samples with weights determined by bias. This leads to recommendations regarding the usefulness of supplementing probability samples with web-based convenience samples. Because the marginal costs of web-based convenience samples are very low, we focus on identifying situations in which the increase in effective sample size (ESS) attributable to the inclusion of the convenience sample may be sufficient to justify a dual-mode approach. We demonstrate that there are limited circumstances under which a supplemental web-based convenience sample may meaningfully improve MSE. While we focus on web-based convenience samples, the discussion that follows applies to other low-cost data collection methods with poor population coverage.

2. Problem context

2.1 Initial conditions

For the combined probability/convenience sample, we propose that the same survey be administered simultaneously to a traditional probability sample (with or

1. Marc Elliott, Ph.D. and Amelia Haviland, Ph.D., Rand Corporation, 1776 Main Street, Santa Monica, CA 90407, U.S.A.

without a web-based response mode) and a web-based convenience sample. We envision a multi-purpose survey with a number of survey outcomes. In this paper we will focus on the estimation of means, but future work might extend these results to other parameters, such as regression parameters. Although we will initially consider cases where the bias of convenience sample estimates is known, we will later consider the extent to which the probability sample provides a means of measuring the unknown bias in each parameter estimate from the convenience sample.

With known bias, one may combine the convenience and probability samples in a manner that minimizes MSE. If estimates from the convenience sample are very biased, the convenience sample will accomplish little. This possibility requires that the probability sample be large enough to stand on its own. Thus, one approach would be to set aside a small portion of the probability sample budget to create a large convenience sample supplement.

For example, consider a survey for which the primary interest is in estimates for the population as a whole, but for which subpopulations estimates would also be desirable if a sample size supporting adequate precision were affordable. Suppose further that one could draw 4,000 probability observations and 10,000 convenience observations for the cost of the probability sample of 5,000. For a given outcome, if bias is large, standard errors increase moderately through a small proportionate loss in sample size; if bias is small overall and within each subpopulation, there might be a “precision windfall,” allowing acceptably precise sub-population analyses.

2.2 Initial bias reduction

We will demonstrate that the bias of convenience sample estimators must be quite small for the sample to be useful, suggesting that it may be best to focus on estimating parameters that are typically subject to less bias than overall unadjusted population estimates of proportions or means, such as regression coefficients (Kish 1985).

Additionally, one might reduce bias by calibrating the convenience sample to known population values (Kalton and Kasprzyk 1986) or by applying propensity score weights that model membership selection between the two samples to observations from the convenience sample (Rosenbaum 2002; Rosenbaum and Rubin 1983). A small set of items can be included to allow the use of either approach. These items might include both items that predict differences between respondents to web surveys and other survey modes, as well as items tailored to the content of the particular survey. The design effect from the resulting variable weights will reduce the ESS for convenience sample estimators, but the low costs of these observations makes compensating for moderate design effects affordable. We then can estimate the remaining bias for a given parameter as the difference between the estimate in the probability sample and the weighted estimate in the convenience sample.

3. Efficiency considerations

3.1 Linear combinations of biased and unbiased estimators of a population mean

The most efficient estimator that is a linear combination of the (weighted) convenience and probability samples is a special case of an estimator given in a result by Rao (2003, pages 57-58). The properties of this estimator lead to general recommendations regarding the conditions of probability sample size, convenience sample size, and convenience sample estimator bias under which the convenience sample meaningfully improves the ESS of the probability sample.

We begin by asking: What is the most efficient estimator of this form when the magnitude of the bias is known? We will later consider relaxing the assumption of known magnitude of the bias.

Let n_1 and n_2 be the effective sample sizes of the probability sample and convenience samples, respectively, after dividing nominal sample sizes by design effects associated with the sample design and non-response adjustments. This includes propensity score or other weighting in the case of the convenience sample. The former population has mean μ, and variance σ_1^2; the latter has mean $\mu + \varepsilon$ and variance σ_2^2, where ε is the known bias remaining after weighting and μ is the unknown parameter of interest. The corresponding sample means have expectation μ and $\mu + \varepsilon$ and variance σ_i^2 / n_i for $i = 1, 2$ under an infinite population sampling model. We assume these two estimators are uncorrelated, as they come from independent samples.

From Rao (2003, pages 57-58), the most efficient composite estimator of μ takes the form

$$\hat{\mu} = \frac{\bar{x}_2(\sigma_1^2 / n_1) + \bar{x}_1(\varepsilon^2 + \sigma_2^2 / n_2)}{\varepsilon^2 + \sigma_1^2 / n_1 + \sigma_2^2 / n_2},$$

with remaining bias

$$\varepsilon_c = \varepsilon\left(\frac{\sigma_1^2 / n_1}{\varepsilon^2 + \sigma_1^2 / n_1 + \sigma_2^2 / n_2}\right)$$

and

$$\text{MSE}_c = \frac{(\sigma_1^2 / n_1)(\varepsilon^2 + \sigma_2^2 / n_2)}{\varepsilon^2 + \sigma_1^2 / n_1 + \sigma_2^2 / n_2}.$$

As can be seen, the composite estimator is a convex combination of the convenience sample and probability sample means. The influence of the former is determined by the ratio of the MSE (here variance) of the probability sample mean to the sum of that term and the MSE of the convenience sample mean. Similarly, the remaining bias is the original bias multiplied by this same ratio, whereas the resultant MSE_c is the product of the two MSEs divided by their sum. Note that bias approaches zero both as $\varepsilon \to 0$ (no selection bias in the convenience sample estimate) and as $\varepsilon \to \infty$ (no weight given to the convenience sample).

3.2 Quantifying the contributions of the convenience sample

We now can evaluate the contributions of the convenience sample based on the known remaining bias in its associated estimators. To this end, we will define several quantities.

Let

$$\mathrm{ESS}_1 = \frac{\sigma_1^2 / n_1}{\mathrm{MSE}_c} n_1 = \left(\frac{\varepsilon^2 + \sigma_1^2 / n_1 + \sigma_2^2 / n_2}{\varepsilon^2 + \sigma_2^2 / n_2} \right) n_1$$

be the effective sample size needed for an unbiased sample mean with the same MSE as the composite estimator. To further simplify this expression, let us define the remaining standardized bias, $E = \varepsilon / \sigma$, and consider the case in which the observations from the convenience and probability populations have equal variance, $(\sigma_1 = \sigma_2 = \sigma)$. In this case, the *increment* to ESS_1 attributable to the convenience sample, the difference between ESS_1 with and without the convenience sample, is

$$\frac{1}{1/n_2 + E^2} = n_2 \left(\frac{1}{1 + n_2 E^2} \right).$$

3.3 Maximum contribution of the convenience sample

As $n_2 \to \infty$, the increment to ESS_1 approaches $1/E^2$. This limit, the inverse of the squared standardized bias, is the maximum possible incremental contribution of the convenience sample to the ESS_1 (abbreviated MICCS). If the MICCS is small, then a convenience sample of any size cannot meaningfully improve MSE. If the MICCS is large enough to be meaningful, we then need to consider what convenience sample sizes are needed to achieve a large proportion of the MICCS.

To develop intuition for the magnitude of E (standardized bias) we consider the important case of a dichotomous outcome, for which $E = \varepsilon / \sqrt{P(1-P)}$ where P is the population probability of the outcome. Table 1 below translates bias for a dichotomous outcome from percentage points to standardized bias and then to the corresponding MICCS for $P = 0.1$ and $P = 0.5$.

Table 1 Maximum contributions of convenience samples to the estimation of a proportion by bias in percentage points

E (Standardized Bias[#])	Overall Prevalence of Outcome 10%	50%	MICCS[&]
	Bias in Percentage Points		
0.01	0.3%	0.5%	10,000
0.02	0.6%	1.0%	2,500
0.05	1.5%	2.5%	400
0.10	3.0%	5.0%	100
0.20	6.0%	10.0%	25

[#] Of estimators of means using only the convenience sample
[&] ESS added with an infinitely large convenience sample relative to no use of a convenience sample.

For a proportion near 50%, a bias of 2.5 percentage points limits the potential increment of ESS_1 to 400. The minimum increment to ESS_1 that offsets the fixed cost of setting up the web-based response mode will vary by user, but we suspect increments of less than 100 will rarely be cost-effective. Table 1 then implies that convenience samples for which the standardized biases of estimators restricted to the convenience sample generally exceed 0.1 standard deviations will rarely prove cost-effective. For a dichotomous variable with P between 0.1 and 0.5 this corresponds to a bias of 3 to 5 percentage points.

How easily are biases of this magnitude achieved with adjusted estimates from convenience samples? Several studies compared propensity-weighted web-based convenience samples to RDD surveys. One (Taylor 2000) advocated the stand-alone use of such convenience samples despite differences of as much as five percentage points in a number of estimates for dichotomous outcomes regarding political attitudes, with standardized bias of 0.05 to 0.10 if one treats RDD as a gold standard. Another (Schonlau, Zapert, Simon, Sanstad, Marcus, Adams, Spranca, Kan, Turner and Berry 2003) does not report magnitudes of differences, but does report that 29 of 37 items regarding health concerns exhibit differences that are statistically significant at $p < 0.01$. Given the reported sample sizes (and optimistically ignoring any DEFF from weighting), it can be shown that significance at that threshold implies point estimates of standardized bias exceeding 0.05 for estimators of 78% of items. The key outcome in a Slovenian comparison of a probability phone sample and a Web-based convenience sample (Vehovar, Manfreda and Batagelj 1999) would be estimated with a standardized bias of more than 0.1 from the convenience sample even after extensive weighting adjustments. It should be noted that there may also be mode effects on responses for the Web mode when compared to a telephone mode among subjects randomized to response mode (Fricker, Galesic, Tourangeau and Yan 2005), so that not all differences between Web convenience samples and non-Web probability samples may result from selection.

3.4 Actual contribution of the convenience sample

While the maximum possible increment (MICCS) is $1/E^2$, the actual increment to ESS_1 can be expressed as $(k/k+1)$ MICCS where $k = n_2E^2$. The shortfall of the actual increment to ESS_1 from the MICCS can then be expressed as $\text{MICCS-}\mathrm{ESS}_1 = 1/[(E^2)(1+n_2E^2)]$. This implies that the returns to ESS_1 diminish with increasing size of the convenience sample, more quickly with large bias since the bias eventually dominates any further variance reduction. Half of the MICCS noted is achieved when the ESS of the convenience sample is equal to the MICCS. For example, if bias is 0.01 standard deviations and a convenience sample has an ESS of 10,000, then the MICCS is 10,000, but the actual incremental contribution to ESS_1 will be 5,000. This suggests that convenience samples with ESS 2-20 times as large as MICCS will suffice for most purposes, which correspond to 67%-95% of the potential gain in ESS. Such heuristics in turn imply collecting 200 - 4,000 such cases when E is relatively large ($E = 0.05$ to 0.10) and 5,000 - 200,000 such cases when E is relatively small ($E = 0.01$ to 0.02). Table 2 provides illustrative examples of the ESS_1 achieved at several combinations of sample sizes and bias.

Table 2 Examples of ESS1 at several sample sizes and levels of standardized bias

n_1 (Probability Sample Size)	n_2 (Convenience Sample Size)	E (Standardized Bias[#])	ESS_1 for the Composite Estimate	ESS_1 / n_1[&]
1,000	1,000	0.01	1,909	1.909
1,000	1,000	0.10	1,091	1.091
1,000	10,000	0.01	6,000	6.000
1,000	10,000	0.10	1,099	1.099
1,000	100,000	0.01	10,091	10.091
1,000	100,000	0.10	1,100	1.100
10,000	1,000	0.01	10,909	1.091
10,000	1,000	0.10	10,091	1.009
10,000	10,000	0.01	15,000	1.500
10,000	10,000	0.10	10,099	1.010
10,000	100,000	0.01	19,091	1.909
10,000	100,000	0.10	10,100	1.010

Number of estimators of means using only the convenience sample

[#] Of estimators of means using only the convenience sample

[&] ESS relative to no use of a convenience sample.

3.5 Precision for estimating bias

Heretofore, we have assumed a known bias in convenience sample estimators; in practice, the bias will need to be estimated using information from both samples. We next explore the extent to which the size of the probability sample also constrains the usefulness of the convenience sample through the need to precisely estimate the remaining bias.

We can estimate ε as the difference between the sample mean of the probability sample and the weighted mean of the convenience sample. The true standard error for the estimate of bias is $\sigma_{\hat{\varepsilon}} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$. If $\sigma_1 = \sigma_2 = \sigma$, the true standard error for the estimate of standardized bias (E) is $\sigma_{\hat{E}} = \sqrt{1/n_1 + 1/n_2}$. No matter how large the convenience sample, this term can never be less than the inverse of the square root of the probability sample size.

It has been demonstrated that the relative error in MSE for a composite estimator is relatively insensitive to small errors in the estimates of bias (Schaible 1978), which is encouraging for well-estimated biases. Unfortunately, unless both the probability and convenience ESS are large, the standard error of the estimate for E is impractically large relative to the values of E that make the convenience supplement useful ($E < 0.10$). For example, suppose that a probability sample of ESS 1,000 and a convenience sample of ESS 5,000 yielded a point estimate of standardized bias of 0.02. If the point estimate were correct, the convenience sample would increase the ESS_1 by 1,667. But this estimate could also have a true bias of 0.088 standard deviations (95% upper confidence limit), which would imply that the increment would be less than 130.

If we assume that the convenience sample size will always be at least twice the probability sample size, these results imply that practical applications of this technique must have a minimum sample size of 1,000-10,000 for the probability sample if they are to address the uncertainty in the magnitude of bias in convenience sample estimators (standard errors of E in the 0.01 to 0.04 range).

4. Discussion

We describe a composite estimator that is a linear combination of an unbiased sample mean estimate from a probability sample and a biased (propensity-score weighted) sample mean estimate from a web-based convenience sample. We use the MSE of this composite estimator to characterize the contributions of the convenience sample to an estimator based only on the probability sample in terms of ESS. We then calculate the maximal contribution of the convenience sample, the role of the convenience sample size in approaching this limit, and the roles of both sample sizes in estimating bias with sufficient precision.

Practitioners sometimes assume that small probability samples are sufficient to estimate the bias in estimates from corresponding convenience samples. Our results suggest otherwise. We demonstrate that the standardized bias of web-based convenience sample estimators after initial adjustments to reduce bias must be quite small (no more than 0.1 standard deviations, and probably less than 0.05 standard deviations) for the MSE of the overall estimate to be meaningfully smaller than it would be without use of the convenience sample. We further demonstrate that convenience sample sizes of thousands or tens of thousands are also needed to realize practical gains. Finally, we demonstrate that a large probability sample size (1,000-10,000) is also needed for reasonably precise estimates of the remaining bias in initially bias-adjusted convenience sample estimators. Because the bias of estimates in an application to a multipurpose survey is likely to vary by outcome, the global decision to substitute a large number of inexpensive surveys for fewer traditional surveys must be made carefully.

The greatest opportunity in cost savings may be in large surveys, simply as a function of their size. On the other hand, the greatest proportionate gains in precision are likely to occur for samples of intermediate size. Gains might also be substantial for large samples in which the main inferences are smaller subgroups. For example, a national survey of 100,000 individuals might make inference to 200 geographic subregions, with samples of 500 for each. If one supplemented this national sample with a very large web-based convenience sample, estimated the bias nationally, and elected to assume that the bias did not vary regionally, one might decrease the MSE of the sub-region estimates substantially through the use of such a composite estimator.

As a final caveat, the conclusions about the limited usefulness of convenience samples with estimator bias of more than 0.1 standard deviations are not limited to attempts to use a composite estimator. The same approach can be applied to show that an estimator based only on a convenience sample of any size with a standardized bias of 0.2 (*e.g.*, ten percentage points for a dichotomous variable with $P = 0.5$) will have an MSE greater than or equal to that of an estimate from a probability sample of size 25.

Acknowledgements

The authors would like to thank John Adams, Ph.D., for his comments on the manuscript, Matthias Schonlau, Ph.D., for his comments on an earlier version of the manuscript, and Colleen Carey, B.A., Kate Sommers-Dawes, B.A., and Michelle Platt, B.A. for their assistance in the preparation of the manuscript.

Marc Elliott is supported in part by the Centers for Disease Control and Prevention (CDC U48/DP000056); the views expressed in this article do not necessarily reflect the views of the Centers for Disease Control and Prevention.

References

Couper, M.P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64, 464-494.

Fricker, S., Galesic, M., Tourangeau R. and Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly*, 69, 370-392.

Kalton, G., and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology*, 12, 1-16.

Kish, L. (1985). *Survey Sampling*. New York: John Wiley & Sons, Inc.

Kypri, K., Stephenson, S. and Langley, J. (2004). Assessment of nonresponse bias in an internet survey of alcohol use. *Alcoholism: Clinical and Experimental Research*, 28, 630-634.

Rao, J.N.K. (2003). *Small Area Estimation*. New York: John Wiley & Sons, Inc.

Rosenbaum, P.R. (2002). *Observational Studies*, 2nd Edition, New York: Springer-Verlag.

Rosenbaum, P.R., and Rubin, D.B. (1983). The central role of the propensity score in observational studies for casual effects. *Biometrika*, 70, 41-55.

Schaible, W.A. (1978). Choosing weights for composite estimators for small area statistics. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 741-746.

Schonlau, M., Zapert, K., Simon, L.P., Sanstad, K., Marcus, S., Adams, J., Spranca, M., Kan, H., Turner, R. and Berry, S. (2003). A comparison between responses from propensity-weighted web survey and an identical RDD survey. *Science Computer Review*, 21, 1-11.

Taylor, H. (2000). Does internet research 'work'? Comparing on-line survey results with telephone surveys. *International Journal of Market Research*, 42, 41-63.

Vehovar, V., Manfreda, K.L. and Batagelj, Z. (1999). Web surveys: Can weighting solve the problem? *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 962-967.