COMPARING NON-PROBABILITY AND PROBABILITY SAMPLE ESTIMATES

Ronaldo Iachan[1], John Boyle[2], Eric Miller[3]
[1]ICF, 530 Gaither Road, Rockville, MD, 20850
[2]ICF, 530 Gaither Road, Rockville, MD, 20850
[3]ICF, 530 Gaither Road, Rockville, MD, 20850

1. Introduction

Local health survey data can help better inform local policy and program implementation than state or national-level data. Probability sample surveys with mail and/or phone components can be slow and costly; moreover, with the advent of new technology, such as call block and caller ID, and declining response rates, they have become increasingly vulnerable to non-response and coverage biases. Against this background, non-probability internet panel surveys have growing appeal for more cost effective procedures for generating timely, actionable data in fit-for-purpose surveys (Hines et al, 2010; Baker et al., 2004).

Internet-based surveys offer unique benefits over more traditional survey methods. The survey administration and data collection process, typically intricate for large scale surveys, is cut down substantially for a variety of reasons. First, the sampling frame is readily available and can easily support the targeting of subpopulations of interest. Second, the online administration expedites data editing (with no scanning or data entry).
Internet panels also offer promise for surveys focused on sensitive topics. Participants are more likely to be honest about their involvement or knowledge of certain activities in a self-administered internet panel survey than in interviewer-administered surveys such as telephone surveys (Hines et al, 2010; Martinsson et al, 2013). These surveys can also ease the effort of the researchers in implementing changes in the survey when new focal points arise. (Harris et al, 2009)

While non-probability samples are potentially more biased than probability samples, a stratified selection approach can reduce the potential for selection bias in internet panel surveys. (Erens et al, 2014). Deep stratification per se allows the matching of multiple demographic and socio-economic dimensions. In addition, a stratified random design can also introduce elements of random selection and support the computation of sampling errors. Multiple studies have compared results obtained from non-probability and probability sample surveys selected from a same or similar populations. A comprehensive evaluation is provided by Yeager et al (2011) with mixed conclusions about the relative accuracy of non-probability internet sample surveys. This article provides an initial assessment of the quality and usefulness of the data at the community level including comparisons to probability sample survey results.

Statistical comparisons with probability samples require measures of variability such as variances, standard errors and confidence intervals. Recent comparisons have focused exclusively on bias (e.g., Yeager et al., 2011, Pew, 2016), and therefore miss one important component of the total error. While these measures of variability are available for the probability sample estimates used as benchmarks, their computation may be more challenging for non-probability samples. Conceptually, variances can be computed with replication methods or with probability sampling framework premised on repeated sampling from the same panel.

This article explores the use of variances for the statistical comparisons between the two samples, a probability sample and a non-probability sample. This approach supports research questions such as: a) do the NPS estimates fall within the confidence intervals computed for the probability sample; b) do the probability sample estimates fall within the confidence intervals computed for the NPS; and c) do the confidence intervals overlap for some, many, or most survey estimates?

2. Methods

2.1 Non-probability samples

.For our investigation, the samples were selected from Research Now® national panels with more than 3 million members in the United States, and more specifically, the e-Rewards panel. Members are included in this panel by invitation from partner organizations, such as airlines, hotels or retailers, and receive small incentives that can then be redeemed for rewards if they participate. Panel members were then sampled and secondarily invited to participate in this online survey. The Research Now® team conducts industry standard quality checks in order to remove panel participants if they consistently provide poor quality or inconsistent data.

Participants were sampled from different communities across the United States; this paper is focused on the Los Angeles County data. Sites were chosen based on the availability of comparison data from the Selected Metropolitan/Micropolitan Area Risk Trends (SMART) Behavioral Risk Factor Surveillance System (BRFSS) data available from the Centers for Disease Control and Prevention (CDC).

Samples were selected based on information provided by panelists in their member profiles, which included the zip code of their residence, age, sex, race, ethnicity, marital status, income and education. Responders received a direct e-mail invitation to the survey (as opposed to internet traffic or routed sample). Panelists received an initial survey invitation, potentially followed by a reminder (no sooner than 36 hours after the initial invite). General subject lines/survey invitation text were used to limit any potential sponsor or topic salience bias. Research Now® continually sampled panelists in an effort to produce a final sample that matches the Census demographics (age group, sex, race/ethnicity, and education level) of each particular community, a standard quota sampling approach for internet-panel surveys. However, the protocol targets the hardest-to-reach populations first (rare and low responding groups in the panel) and then sends additional invitations as quotas are attained, with updated demographic requirements based on prior response.

The Los Angeles County experimental study included two arms. The first arm was selected using the usual approach adopted in internet panel samples. This method, in essence a quota sampling approach also used in our previous experiments, is described briefly above and in more detail elsewhere (Iachan et al, 2015, Iachan et al., 2016). The second arm was selected using a stratified random sampling method. The initial sample distribution matched the local demographics for LA County along key characteristics (age group, sex, race/ethnicity, and education level). Unlike the quota sampling approach, no further attempts to control the sample distribution were made after selection. The second arm also followed more rigorous follow-ups to minimize non-response and the associated biases.

For the non-probability sample arms, variances can be estimated using super-population approaches or replication methods. A cruder approximation may be based on simple random sampling formulas, an approximation that seems more valid for the stratified random sampling arm than for the quota sampling arm. These variances incorporate measures of the effects of unequal weighting for the different methods used in the non-probability samples.

2.2 Weighting the panel data

Potential biases can be greatly reduced with the use of statistical survey weighting techniques which aim at controlling for potential selection biases by using a range of covariates in the adjustments. Raking methods have been used in several studies (e.g., Iachan et al., 2015) to allow for deeper post-stratification for internet panel survey data. Weighting techniques based on propensity adjustments for internet survey data have been considered and compared with post-stratification adjustments in several studies (Dever et al, 2008; Lee, 2006; Lee and Valliant, 2009; Loosveldt et al, 2008; Lensvelt-Mulders et al, 2009). While the panel population is necessarily skewed to some degree, the samples were balanced so that the respondents (unweighted) distribution is close to the population demographics in each community. The raking approach for this sample used 2013 data from the American Community Survey (ACS) for the following weighting variables: age, gender, race/ethnicity, education and marital status.

2.3 Comparing non-probability and probability samples

We compared results from the three sites chosen from the internet panel with data from a probability sample survey from the same or similar communities, using SMART BRFSS data. The BRFSS is an annual telephone survey, with a combined landline/cell phone sampling frame, conducted by the CDC in all 50 states. BRFSS data for larger cities and counties are available through the so-called SMART BRFSS. The most recent SMART BRFSS data available for comparisons were from 2012.

We used county-level SMART BRFSS data from Los Angeles County. Metropolitan and Micropolitan Statistical Areas (MMSA) SMART BRFSS data was necessary for the comparisons.
(http://www.cdc.gov/brfss/smart/2012/2012_smart_brfss_mmsa_methodology.pdf).
There were slight variations in question wording between internet-panel survey and SMART BRFSS. Appendix 1 presents the question wording (or combination of questions) used from each survey and topic.

We computed weighted estimates and the associated variance estimates for the local, Los Angeles County (SMART) BRFSS as well as for the panel survey data. The analysis included a range of estimates based on questions which were comparable across the two surveys. Overall weighted estimates of each health indicator from the internet-panel sample and SMART BRFSS were compared at the county level. The standard errors and 95% confidence intervals for the SMART BRFSS accounted for the complex sampling design and weighting effects. To assess the comparability between community-level estimates based on probability and non-probability samples, we calculated the difference in point estimates and determined how often point estimates from the internet-panel data fell within the confidence intervals of SMART BRFSS. Conversely, we assessed how often point estimates from BRFSS fell within the confidence intervals for the panel sample estimates. All analyses were conducted using SAS Survey Procedures (9.4).

3. Results

This section illustrates the comparisons with a few key estimates. Figures 1-4 show the confidence intervals (CI's) computed for the two arms of the LA panel study side by side with the BRFSS CI's for four health indicators: a) asthma prevalence, b) diabetes prevalence, c) cigarette smoking prevalence, and d) categories for body mass index (BMI). For the latter variable, the 4 categories distinguish underweight, normal weight, overweight and obese persons,
The figures show that the CIs tend to overlap for the probability sample and the non-probability sample arms. Not surprisingly, variances tend to be larger for the arm based on a stratified random sample due to unequal weighting effects; for this sample, the weights need to work harder to make the weighted sample resemble the population distribution. For the other arm, this population matching is enforced by the quota sampling approach.

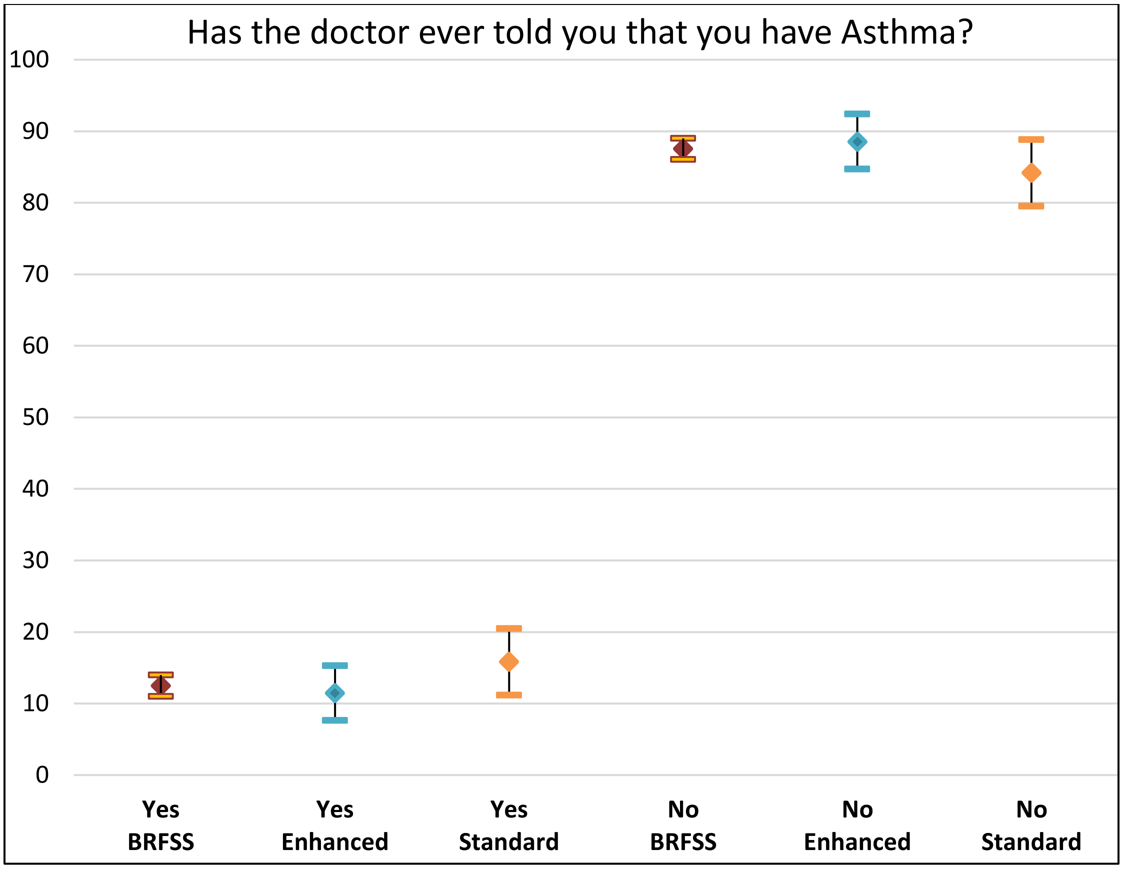

Figure 1: Confidence Intervals for Asthma

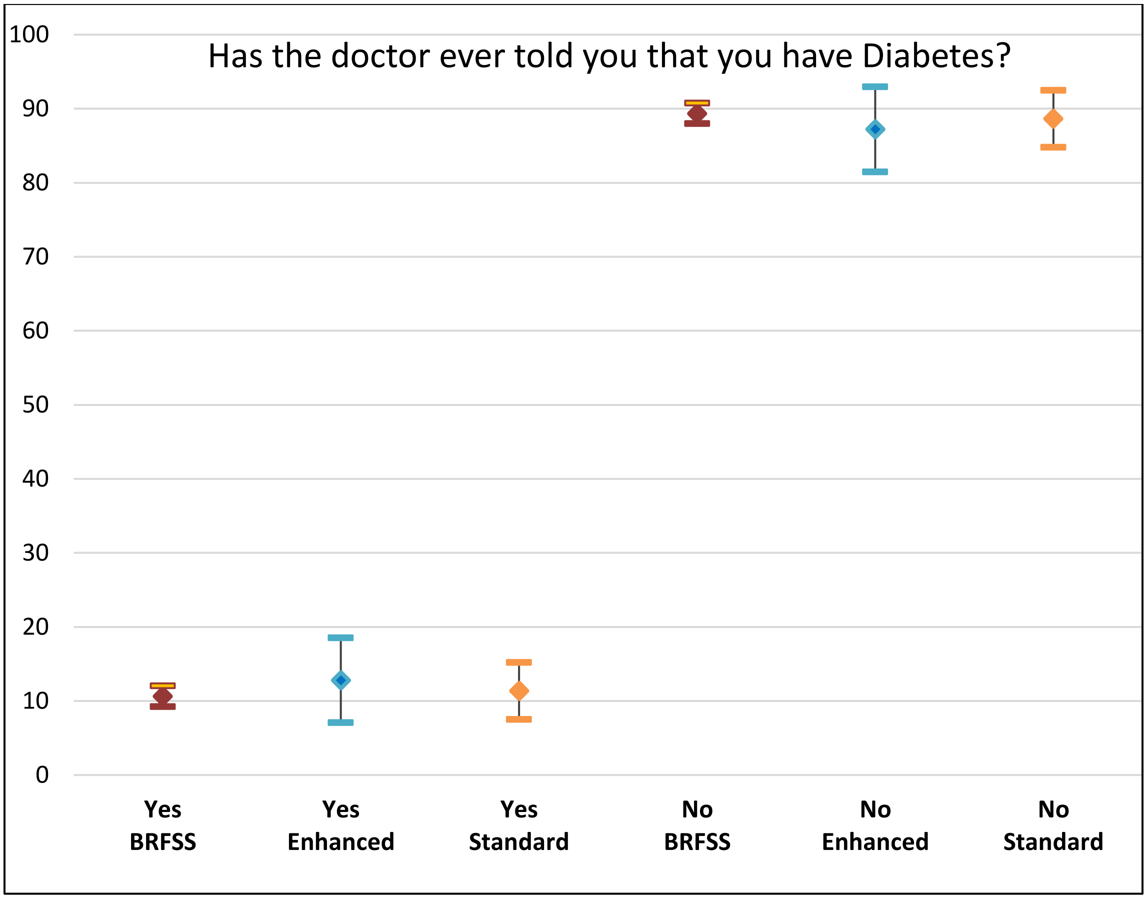

Figure 2: Confidence Intervals for Diabetes

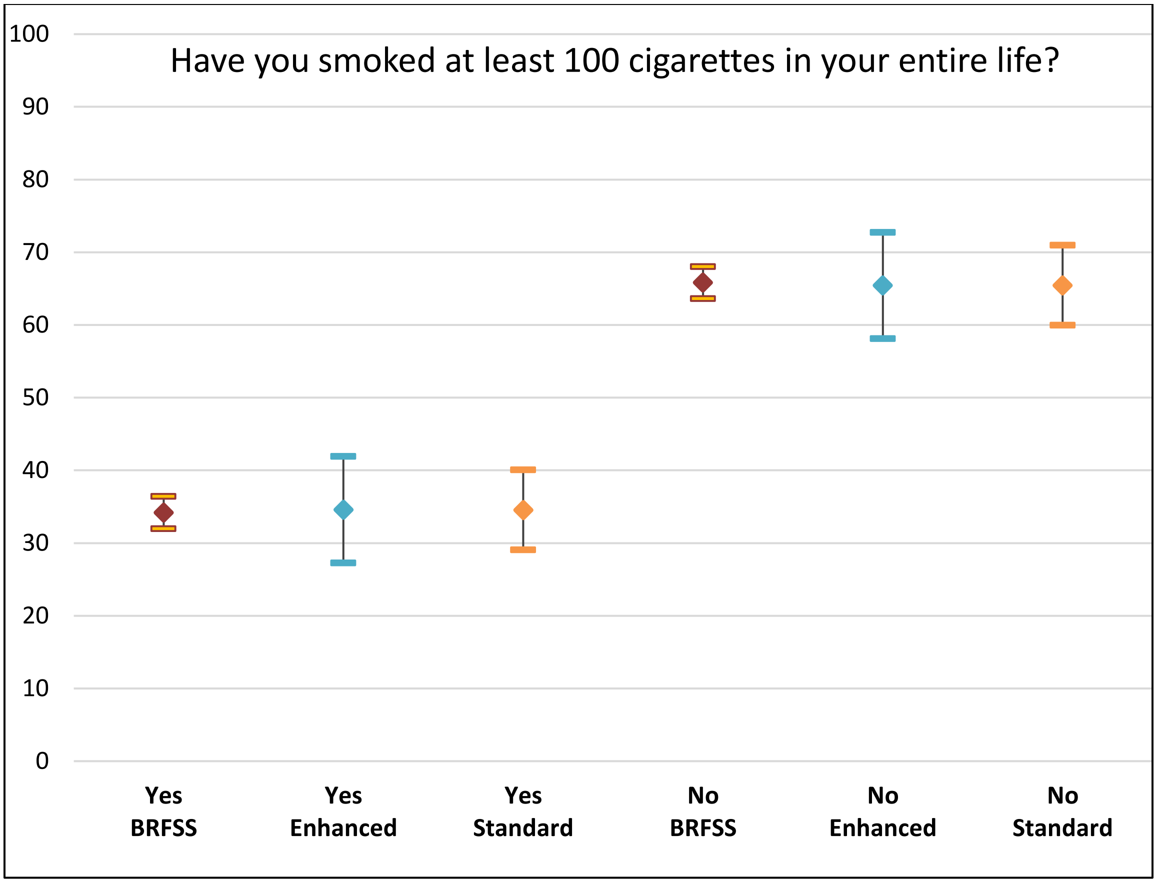

Figure 3: Confidence Intervals for Cigarette Smoking

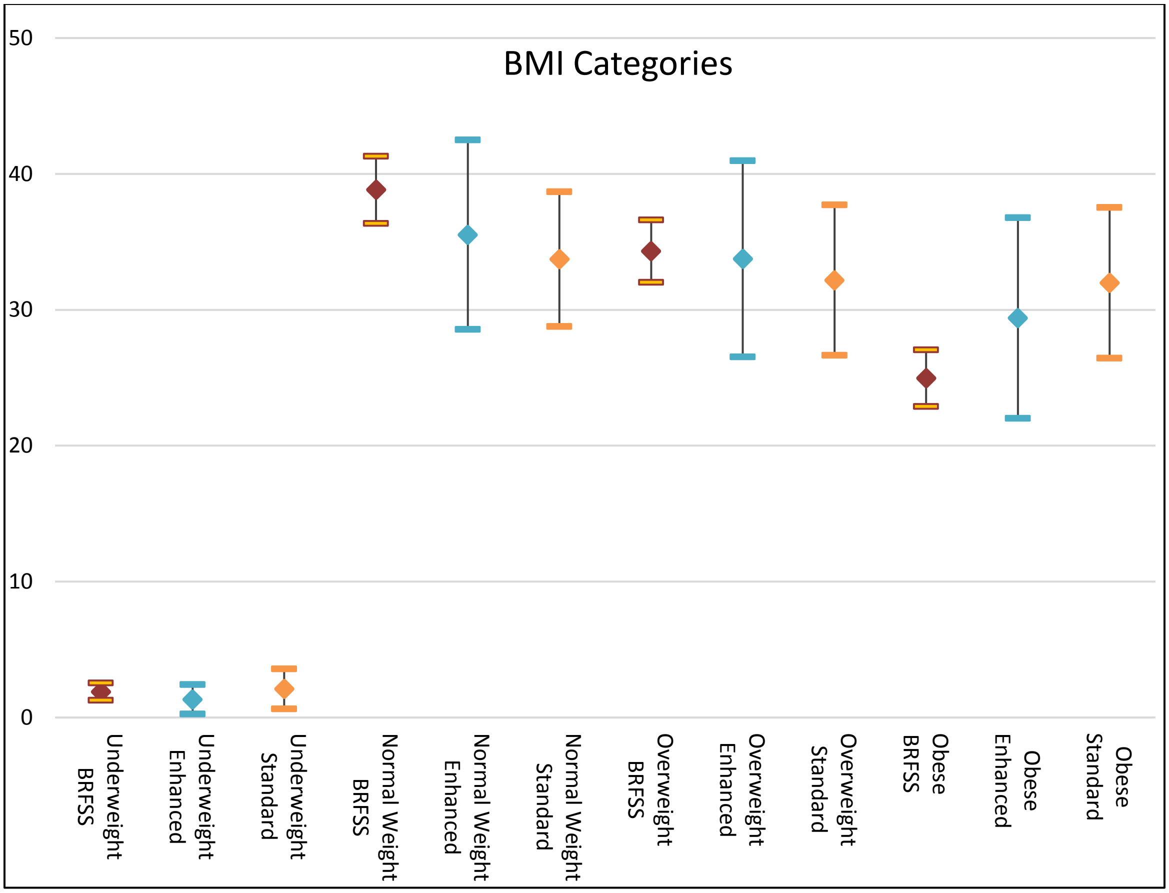

Figure 4: Confidence Intervals for BMI Categories

4. Conclusions

This analysis adds to the current base of knowledge on nonprobability samples by incorporating the use of variance in comparison with probability-based sampling data from similar geographic areas. Variance estimates from the non-probability data were calculated by direct estimation methods which may be particularly accurate for the arm sample selected with stratified random sampling. With two types of sampling processes in place with the internet-panel data, we were also able to compare the variance levels with the different methods.

Overall, we found substantial overlap between estimates of probability- and non-probability-sampled data. While there may still be valid concerns about accuracy and bias in non-probability samples, these results provide some reassurance that non-probability data can have a role in health surveys and fit-for-purpose use. Additionally, use of variance estimates provides another method to evaluate non-probability samples against probability samples.

References

Baker, R., M. Brick, N. Bates, M. Battaglia, M. Couper, J. Deaver, K. Gile, and R. Tourangeau. 2013. Summary report of the AAPOR task force on non-probability sampling. Journal of Survey Statistics and Methodology, 3(4).

Baker, R., D. Zahs, and G. Papa. 2004. "Health Surveys in the 21st Century: Telephone vs. Web." Presented at the Eighth Conference on Health Survey Research Methods, Peachtree City, Georgia, 143-148

Bethlehem, J. and S. Biffignandi. 2010. "Selection Bias in Web Surveys." International Statistical Review. 78, 161-188.

Boyle, J., R. Iachan, L. Berman and J. Dayton. 2015. "Fit for Purpose Community Health Surveys: An Experiment in Three Communities." Presented at the American Association for Public Opinion Research. May 2015. Hollywood, FL.

Boyle, J., T. Fakhouri, N. Freedner-Maguire, and R. Iachan. 2016. "Characteristics of the Population of Internet Panel Members." Submitted for publication.

Brick, M., N. Mathiowetz, S. Cho, S. Cohen, R. Igielnik, S. Keeter, and K. McGeeney. 2015. "Weighting and Sample Matching Effects for an Online Sample." Presented at the American Association for Public Opinion Research. May 2015. Hollywood, FL.

Chang, L. and J. Krosnick, J. 2009. "National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality." Public Opinion Quarterly. 73, 641-678.

Couper, M. 2007. "Issues of Representation in eHealth Research (with a Focus on Web Surveys)." American Journal of Preventive Medicine, 32, S83-S89.

Dever, J., A. Rafferty, and R.Valliant. 2008. "Internet Surveys: Can Statistical Adjustment Eliminate Coverage Bias?" *Survey Research Methods*, 2, 47-62.

Diamond, A. and J. Sekhon. 2012. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics*.

Duffy, B., K. Smith, G. Terhanian, and J. Bremer. 2005. "Comparing Data from Online and Face-to-Face Surveys." International Journal of Market Research, 47, 47-62.

Erens, R. S. Burkill, M. Couper, F. Conrad, S. Clifton, C. Tanton, A. Phelps, C. Mercer, P. Sonnenberg, P. Prah, K. Mitchell, K. Wellings, A. Johnson, and A. Copas. 2014. "Nonprobability web surveys to measure sexual behaviors and attitudes in the general population: A comparison with a probability sample interview survey." Journal Medicine Internet Res., 16(12).

Harris, K., M. Schonlau, and N. Lurie. 2009. "Surveying a nationally representative internet-based panel to obtain timely estimates of influenza vaccination rates." Vaccine 27, no. 6, 815-818.

Hines, D., E. Douglas, and S. Mahmood. 2010. "The effects of survey administration on disclosure rates to sensitive items among men: A comparison of an internet panel sample with a RDD telephone sample." Computers in Human Behavior, 26, 1327-1335.

Iachan, R. 2010. "A new iterative method for weight trimming and raking". Paper presented at the Joint Statistical Meetings, Vancouver, Canada.

Iachan, R., J. Boyle, and T. Fakhouri. 2015. "Weighting Web Panel Data at the Community Level." Presented at the American Association for Public Opinion Research meetings, Hollywood, FL.

Isaksson, A. and L. Sunghee. 2005. "Simple Approaches to Estimating the Variance of the Propensity Score Weighted Estimator Applied on Volunteer Panel Web Survey Data – a Comparative Study." UCLA Center for Health Policy Research. 3143-3149.

Lee, S. 2006. "Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys." *Journal of Official Statistics*. 22, 329-349.

Lee, S. 2006. "An Evaluation of Nonresponse and Coverage Errors in a Web Panel Survey." Social Science Computer Review, 2, 460–75.

Lee, S. and R. Valliant. 2009. "Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment." *Sociological Methods Research*. 37, 319-343.

Lee, B., J. Lessler, and Stuart, E. 2010. "Improving propensity score weighting using machine learning." *Statistics in Medicine*, 29: 337-346.

Martinsson, J., S. Dahlberg and S. Lundmark. 2013. "Is accuracy only for probability samples? Comparing probability and non-probability samples in a country with almost full internet coverage." Presented at the Midwest Political Science Association.

Lensvelt-Mulders, G., P. Lugtig and M. Hubregtse. 2009. "Separating Selection Bias and Non-coverage in Internet Panels Using Propensity Matching." *Survey Practice*, 2.

Rivers, D. 2007. "Sample Matching for Web Surveys: Theory and Application." Paper presented at the Joint Statistical Meetings, Salt Lake City, Utah.

Rubin, D. 1979. "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies." Journal of the American Statistical Association. 74, 318-328.

Scherpenzeel, A. 2011. "Data collection in a probability-based internet panel: How the LISS panel was built and how it can be used." Sociologique, 109, 56-61.

Schonlau, M., K. Zapert, L. Simon, K. Sanstad, S. Marcus, J. Adams, M. Spranca, H. Kan, R. Turner and S. Berry. 2004. "A Comparison between Responses from a Propensity-Weighted Web Survey and an Identical RDD Survey." Social Science Computer Review 22, 128–138.

Terhanian, G. and J. Bremer. 2012. "A Smarter Way to Select Respondents for Surveys?" International Journal of Market Research. 54, 751-780.

Yeager, D., J. Krosnick, L. Chang, H. Javitz, M. Levendusky, A. Simpser and R. Wang. 2011. "Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples." Public Opinion Quarterly, 20, 709-747.