Proceedings of Statistics Canada Symposium 2018
Combine to Conquer: Innovations in the Use of Multiple Sources of Data

The Utility of Using Web Surveys to Provide Official Estimates for Major Health Outcomes, A Pilot Study

Yulei He, Hee-Choon Shin, Bill Cai, Vladislav Beresovsky, Paul Scanlon, Van Parsons, Peter Meyer, and Jennifer Parker[1]

Abstract

Given the spiking cost and resource constraints of traditional surveys, web surveys are being increasingly used as a less-costly alternative. However, past literature has well documented general limitations of web surveys. Correspondingly, there also exists research on understanding and improving their utilities. From a statistical health agency's perspective, one major unanswered question is whether web surveys can be used to produce national, official estimates of major health outcomes (e.g. diabetes prevalence, insurance coverage). Therefore, the National Center for Health Statistics (NCHS) is conducting a series of web survey studies. The two-fold aim of these studies is to assess the utility of web surveys for producing official estimates of health items and to evaluate their use for examining question response patterns. These studies also include data from the National Health Interview Survey (NHIS), which is a major data source for producing official health estimates for the US general population. This paper presents the background of this project and some initial comparisons between estimates from the web survey and the NHIS. The results showed that the weighted estimates of the health outcomes from the web sample tend to be worse than those from the NHIS. The apparent differences of the estimates between the web survey and NHIS raise some concerns with the use of web surveys to replace traditional surveys in producing official estimates for health outcomes in general.

Key Words: Health Survey; National Health Interview Survey; Official Statistics; Probability Panel; Web Survey.

1. Introduction

Sample surveys are a major approach to collecting information from human subjects for scientific research purposes. In the past, the survey approach has mainly relied on three methods of data collection: face-to-face interviews, telephone interviews, and mail surveys. For many years, these traditional survey methods have been effective in collecting health-related information and have provided national official estimates of health conditions. Successful examples include the National Health Interview Survey (NHIS), which is based on face-to-face interviews and conducted by National Center for Health Statistics (NCHS). NHIS is a principal data source for providing official statistics of important health outcomes such as health insurance coverage and prevalence of major diseases (e.g., diabetes). Additional details can be found, for example, in http://www.cdc.gov/nchs/nhis.htm.

However, all major surveys using traditional data collection methods are impacted by decreasing response rates (Williams and Brick 2017). One of the main reasons for this decrease could be the increasing difficulty to contact sample units and to convince them to participate surveys (Singer 2006). Increasing nonresponse rates would naturally bring up the concern of nonresponse bias of the survey estimates (Tourangeau 2017). In addition, for face-to-face, large-scale health surveys, the increasing unit cost creates additional challenges. Given the increasing lack of resources and decreasing response rates, survey researchers and administers are compelled to find an economical alternative approach to collecting information on the nation's health status while retaining the scientific and methodological rigor of probability surveys.

In the past 20 years or so, the field of survey methodology has been experiencing an innovative and challenging expansion – web surveys (e.g. Tourangeau et al. (2013), Baker et al. (2013) and references therein.) The development of the internet and web has fundamentally changed the structure of our daily communication channels. Exchanging electronic mail (e-mails) and sending instant messages over the web are now regarded as ordinary activities in most

[1]Yulei He, National Center for Health Statistics, US Centers for Disease Control and Prevention, 3311 Toledo Road, Hyattsville, MD, USA, 20782 (wdq7@cdc.gov)

developed countries including the US. Taking advantage of e-mail and web-based technology to develop, administer, and implement web surveys appears to be a natural choice. Not only do web surveys represent an advance in the evolution of self-administered questionnaires, the overall cost in web surveys is significantly lower than traditional data collection methods such as face-to-face interviews, in which each completed case may require contact with an interviewer and thus entails high interviewer payments and logistical costs.

In private industry, web surveys have been readily applied in marketing business or opinion polls. Web surveys have been frequently used in business or establishment surveys, replacing or supplementing mail surveys. Their usage is also increasing in government surveys. They are increasingly used in population censuses (e.g., web surveys are being used to as a mode of data collection in the American Community Survey conducted by the U.S. Census). Statistical agencies have explored using web surveys as part of mixed-mode strategy. For example, in 2013 NCHS carried out a telephone/web follow-up survey for a subset of 2012 NHIS sample adult respondents.

However, despite their budgetary appeal and potential convenience, existing research has documented inherent drawbacks of web surveys that might lead to various forms of errors. One major concern is that the low cost of web surveys often reflects a nearly complete disregard of the probability sampling principles. Yet these probability sampling principles are the backbones behind most high-quality surveys (e.g., NHIS) that are designed to make inferences to broad populations and to produce official statistics by statistical agencies. In other words, many web surveys are based on convenience samples, and thus it is often unclear whether the sample drawn from the web surveys is representative of the target population, further casting a doubt on the accuracy of the web-survey based estimates. This observation is based on a large body of literature of comparing estimates from web surveys with those from traditional probability surveys (e.g., Lee and Valiant 2009; Bethlehem 2010).

More specifically, most of the web surveys in practical use were based on volunteer (self-selected/opt-in) panels, which arguably lack the representativeness of the conventional probability samples (Couper 2000). In spite of the general concern about the representativeness of opt-in web surveys, there has been extensive research on the performance of using these web surveys for measuring health items. Couper et al. (2007) listed several research projects using internet surveys to gather information on various health-related topics across multiple clinical areas. More recent literature can be found, for example, in Erens et al. (2014) and Nelson et al. (2014). These studies were also targeted to specific health measures and often conducted on small-scales. The opt-in nature of these web survey studies as well as their limited scopes have posted certain limitations to the study conclusions and make them less generalizable.

As the U.S.'s primary statistical health agency, NCHS is interested in the feasibility of utilizing web surveys in providing national official statistics of a wide spectrum of health outcomes. This is a very broad research problem, and thus we have been conducting a series of web survey studies for investigation. The specific goal of this paper is to describe the background and implementation of the first web survey study and provide some descriptive statistics. The estimates of the web survey are compared with those of NHIS to provide some insight on the feasibility of using them for official estimates. The rest of the paper is organized as follows. Section 2 describes the design and implementation of the web survey. Section 3 presents the estimates and comparison study with NHIS. Section 4 summarizes the findings and lays out future research plans.

2. Study Design

In this paper, we refer to the web survey conducted as the Research and Development Survey (RANDS). Our study includes describing and comparing the health estimates from RANDS with corresponding health estimates from the NHIS, which provides a reliable snapshot of the health status of U.S. general population.

2.1 Gallup Online Panel

In general, web survey data can be collected through either probability sampling or non-probability sampling. Couper (2000) classified web surveys into multiple types depending on how the samples are collected. However, there is a consensus in the existing literature that estimates from non-probability surveys based on volunteer opt-in panels might

not have desirable statistical properties (Baker et al. 2013). That is, their population representativeness would be doubtful. This led to only including probability panel-based web surveys in this project.

The sample data were collected using the Gallup Panel internet surveys. The Gallup Panel is a research panel that is representative of the entire U.S. population. More specifically, at the time of the data collection for RANDS, the Gallup Panel selected potential members using random-digit-dialing (RDD) of landline telephones and cellphones or address-based sampling (ABS) to contact U.S. households at random. Once Gallup recruited a person to join the Panel, depending on his/her Internet access status, Gallup contacted him or her to complete surveys via either email or mail. Gallup also contacted members by phone. The ability to harness the power of multiple survey modes to target a specific population is not achievable in any opt-in panels. Details of the Gallup panel used for RANDS can be found in Rookey et al. (2008).

2.2 Sampling Design and Data Collection

The 1st wave of RANDS was originally designed to have a sample size of 2000 with a 40% response rate. Random sampling was used in multiple strata defined by race/ethnicity (Non-Hispanic White-Only, Non-Hispanic Black-Only, Non-Hispanic Other, Hispanic), age (18-34, 35-54, 55+), and education (High School or Less, Some College, College Graduates). The 1st wave data collection started on 11/02/2015 and ended on 12/09/2015. Because of the lower-than-expected response rate, multiple reminders were sent to the survey participants. Overall, a total of 9,809 respondents were invited to complete the survey. When the survey was pulled from the field on 12/09/2015, a total of 2,304 had completed the survey, for an overall response rate of 23.5%. More specific information can be found in (Gallup fields work reference).

2.3 Questionaire Design

To allow for comparison between the NHIS and the RANDS, selected questions from NHIS were used in the RANDS questionnaire. After consulting with subject-matter experts, a sub-set of 88 NHIS questions (primarily from the NHIS sample adult questionnaire were used. Detailed information about questions used in RANDS is available upon request.

3. Data Analysis

Gallup Inc. provided NCHS survey weights for RANDS data that took into account the overall panel composition and the survey response using post-stratification and nonresponse adjustment. The RANDS survey weights were normalized so that the weighted sample size would be equal to the unweighted sample size of respondents (n=2304). Throughout this paper, we treat RANDS as a probability survey and use their survey weights for estimation.

3.1 Using NHIS 2015 Data for Comparison

The main comparison data (i.e., the reference survey) were constructed from NHIS. More specifically, a subset of NHIS survey collected from October-December, 2015 was extracted (n=7723). The corresponding NHIS survey weights are targeted for providing annual estimates.

3.2 Demographics

Throughout this paper, the statistical comparison is based on the survey-data adjusted Chi-square tests (i.e., Rao-Scott tests Scott and Rao 1981). All the estimates are based on observed data because the item nonresponse rates are fairly small (e.g., less than 5%) in general. The income variable has high item nonresponse for both surveys.

Table 3.2-1 shows the weighted distributions of several demographic variables in RANDS and NHIS. There were differences according to education, income, marital status, and race-ethnicity. RANDS estimates show higher

percentages for less than high school or high school education. RANDS also has more nonresponses for the income question. Most noticeably, compared to the NHIS estimates, the RANDS estimates indicate a higher percentage of Non-Hispanic white adults and a lower percentage of Non-Hispanic Asian and other groups.

Table 3.2-1
Distribution of Demographic Groups in RANDS and NHIS, Wave I

	RANDS		NHIS	
Variable	Mean	S.E.	Mean	S.E.
Age				
18-36 yrs	29.5%	0.91%	29.6%	0.87%
35-55 yrs	33.7%	0.99%	34.4%	0.82%
55-65 yrs	17.0%	0.92%	16.7%	0.59%
65-75 yrs	13.4%	0.85%	11.3%	0.49%
> 75 yrs	6.4%	0.62%	8.0%	0.41%
Gender				
Female	51.1%	1.49%	51.8%	0.79%
Male	48.9%	1.49%	48.2%	0.79%
Census Region				
Northeast	18.1%	1.18%	18.6%	0.84%
Midwest	21.3%	1.20%	22.4%	0.89%
South	36.8%	1.44%	36.8%	1.01%
West	23.8%	1.22%	22.2%	0.92%
Education*				
Less than High School	3.6%	0.69%	15.3%	0.66%
High School	37.0%	1.05%	20.9%	0.73%
Associate, some college	29.2%	0.92%	30.4%	0.76%
Bachelor or higher degree	30.3%	0.90%	33.4%	0.96%
Income*				
Missing	22.9%	1.28%	8.6%	0.48%
< $50,000	29.5%	1.37%	41.0%	0.91%
$50,000-$99,999	26.4%	1.26%	26.5%	0.75%
> $100,000	21.2%	1.16%	23.9%	0.88%
Marital Status*				
Married/With Partner	62.4%	1.46%	60.4%	0.84%
Never Married	23.2%	1.25%	22.1%	0.74%
Separated/Divorce/Widow	14.4%	1.03%	17.5%	0.57%
Racial-ethnic groups*				
Non-Hispanic White	72.4%	0.93%	65.3%	1.00%
Non-Hispanic Black	11.5%	0.71%	12.1%	0.62%
Non-Hispanic Asian	1.0%	0.20%	5.9%	0.35%
Hispanic	14.8%	0.74%	15.7%	0.76%
Other	0.3%	0.08%	1.1%	0.18%

Notes: S.E. stands for standard error; * denotes significant at 5% level from the chi-square test comparing RANDS and NHIS.

3.3 Major Health Indicators

Table 3.3-1 presents some descriptive statistics of selected variables that are commonly used to summarize access to health care, health service usage, health conditions, and related health behavior. There exist significant differences between the RANDS estimates and NHIS estimates for some of the major health indicators. Although the estimates from RANDS web survey participants indicated high percentages of adults covered by health insurance (over 90%) and heavily using computers for health-related information (more than 80%), estimates from RANDS for problems in getting affordable and timely health care services were higher compared with those from the NHIS. RANDS also led to higher estimates than the NHIS for items such as having obesity, having more borderline diabetes, asthma, and lung diseases, feeling more nervous and emotional problems, and being more worried about some of the basic life support such as food.

Table 3.3-1
Distribution of Major Health Indicators in RANDS and NHIS, Wave I

	RANDS		NHIS	
Variable	Mean	S.E.	Mean	S.E.
Are you covered by any kind of health insurance or some other kind of health plans?*	93.4%	0.80%	90.5%	0.51%
During the past 12 months, was there any time when you couldn't afford and didn't get any of the health care services?*	31.5%	1.41%	15.9%	0.57%
Have you delayed getting care for any reason in the past 12 months?	26.1%	1.32%	10.8%	0.44%
Obesity*				
Underweight (BMI < 18.5)	1.2%	0.35%	1.7%	0.19%
Normal weight (BMI >=18.5)	30.9%	1.40%	34.9%	0.74%
Overweight (BMI >=25)	33.8%	1.43%	33.9%	0.67%
Obese (BMI >=30)	34.1%	1.41%	29.5%	0.77%
Other than during pregnancy, have you ever been told by a doctor that you have diabetes or sugar diabetes?*				
No	81.2%	1.12%	85.6%	0.51%
Borderline	9.8%	0.89%	5.3%	0.35%
Yes	9.0%	0.79%	9.1%	0.42%
Have you been told by a doctor that you have hypertension (high blood pressure)?	32.0%	1.26%	30.6%	0.75%
Have you even been told by a doctor that you have asthma?*	17.2%	1.14%	12.9%	0.53%
Do you have any one of the emphysema, COPD, or chronic bronchitis?*	9.5%	0.89%	6.2%	0.39%
How often do you now smoke cigarettes? Every day, some days, or not at all?				
Not at all:	87.3%	0.99%	85.7%	0.55%
Sometimes	4.1%	0.59%	3.6%	0.26%
Everyday	8.6%	0.84%	10.7%	0.52%
In any one year, have you had at least 12 drinks of any type of alcoholic beverage?*	70.3%	1.37%	64.4%	0.81%

During the past 30 days, did you feel all or most of time any of sad/nervous/restless/hopeless/no effort/worthless*	18.4%	1.19%	13.5%	0.60%
How often (times per week) do you do any combination of the leisure time activities? (Scale 0-7)*	3.8190	0.0732	2.9496	0.0497
Self-rated health status*				
Excellent health	12.7%	0.99%	28.1%	0.70%
Very good	41.5%	1.48%	31.6%	0.76%
Good	34.2%	1.40%	27.2%	0.70%
Fair	9.6%	0.87%	10.2%	0.49%
Poor	2.0%	0.41%	2.9%	0.23%
I worried whether my food would run out before I got money to buy more*				
Often	5.0%	0.71%	4.0%	0.31%
Sometime	15.8%	1.12%	9.8%	0.46%
Never	79.3%	1.24%	86.1%	0.56%
Working for pay at a job or business?	60.4%	1.37%	60.2%	0.83%
During the past 12 months, have you ever used computers to look up health information on the internet?*	82.8%	1.14%	52.9%	0.90%

Notes: S.E. stands for standard error. * denotes significant at 5% level from the chi-square test comparing RANDS and NHIS.

3.4 Other Variables

For the brevity of the presentation, we do not go over the estimates of all the survey questions. They are available upon request. The major health indicators shown in Table 3.3.1 acted as the gateway or summary questions in the survey. Estimates from corresponding follow-up questions in general showed similar comparative patterns, yet provided finer details. Overall, compared with the estimates from NHIS, estimates from RANDS would indicate that adults are less healthy, have more emotional issues, are engaged in more risky behaviors such as drinking, and are have more issues to getting access to affordable and timely medical care.

4. Conclusion

Based on our knowledge, RANDS is perhaps the very first use of a probability panel-based web survey for a comprehensive list of health indicators led by a statistical agency, despite extensive studies on the use of opt-in panel-based web surveys. This paper describes some of the background information and provides descriptive analyses of RANDS wave 1 data. Our study show that web surveys can yield comparable estimates with the traditional face-to-face survey (i.e., NHIS) for some demographic and health items (e.g., hypertension). However, for many other key health items, the direct estimates from the web survey are considerably different from those of the traditional survey. This finding raises concerns about the feasibility of using direct estimates from probability panel-based web surveys as official statistics of general health outcomes.

Major reasons for the apparent difference might include sampling error and coverage error. Although we intentionally choose to use the Gallup panel, which is a probability panel generated by RDD, its coverage properties are less clear compared with the well-established national surveys such as NHIS. The high nonresponse rate (more than 70% in general) of RANDS might also bias the estimates. We note that Gallup Inc. used some traditional post-stratification and nonresponse adjustments by certain demographic groups. The next step of our research is to investigate the

feasibility of using state-of-the-art modeling techniques to calibrate web survey estimates. Some of the related literature for calibration can be found in Lee (2006a, 2006b) and Lee and Valliant (2009). Although developed for non-probability data, we expect that these and related approaches for calibrating to a 'gold standard' will be directly applicable to a variety of data calibrated to the NHIS. In our context, since our interest focuses on health outcomes, the control-totals needed in common calibration methods can go beyond the typical demographics and be extended to health variables, which can be readily estimated from NHIS, the reference survey.

References

Baker R. P., J. M. Brick, N. A. Bates, M. Battaglia, M. J. Couper, J. A. Dever, K. J. Gile, and R. Tourangeau (2013), "Summary Report of AAPOR Task Force on Non-probability Sampling", *Journal of Survey Statistics and Methodology*, 1, pp. 90-143.

Bethlehem J. (2010), "Selection Bias in Web Surveys", *International Statistical Review*, 78, pp. 161-188.

Cantrell M. A., and P. Lupinacci (2007), "Methodological Issues in Online Data Collection", *Journal of Advanced Nursing*, 60, pp. 544-549.

Couper M. P. (2000), "Web Surveys: A Review of Issues and Approaches", *Public Opinion Quarterly*, 64, pp. 464-494.

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

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

Erens B., S. Burkill, M. P. Couper, F. Conard, S. Clifton, C. Tanton, A. Phelps, J. Datta, C. H. Mercer, P. Sonnenberg, P. Prah, K. R. Mitchell, K. Wellings, A. M. Johnson, and A. J. Copas (2014), "Nonprobability Web Surveys to Measure Sexual Behaviors and Attitudes in the General Population: A Comparison with a Probability Sample Interview Survey", *Journal of Medical Internet Research*, 16, p. e276.

Gallup Inc. (2018), *CDC/NCHS/RANDS Field Work Summary*.

Greby S. (2014), *Surveys to Monitor Influenza Coverage among Pregnant Women Final Report*, Abt Associates. Contract # GS-10F-0086K.

Lee S. (2006a), "An Evaluation of Nonresponse and Coverage Errors in a Prerecruited Probability Web Panel Survey", *Social Science Computer Review*, 24, pp. 460-475.

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

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

McCabe S. E., C. J. Boyd, M. P. Couper, D. Crawford, and H. D'Arcy (2002), "Mode Effects for Collecting Alcohol and Other Drug Use Data: Web and US Mail", *Journal of Studies on Alcohol and Drugs*, 63, pp. 755-761.

Nelson E.J., J. Hughes, J. M. Oakes, J. S. Pankow, and S. L. Kulasingam (2014), "Estimation of Geographic Variation in Human Papillomavirus Vaccine Uptake in Men and Women: An Online Survey Using Facebook Recruitment", *Journal of Medical Internet Research*, 16, p. e198.

Rookey B.D., S. Hanway, and D. A. Dillman (2008), "Does a Probability-Based Household Panel Benefit from Assignment to Postal response as an Alternative to Internet-Only?", *Public Opinion Quarterly*, 72, pp. 962-984.

Scanlon P. (2018), *Using Targeted Embedded Probes to Quantify Cognitive Interviewing Findings*.

Schonlau M., A. van Soest, A. Kapteyn, and M. P. Couper (2009), "Selection Bias in Web Surveys and the Use of Propensity Score", *Sociological Methods and Research*, 37, pp. 291-318.

Scott A.J., and J. N. K. Rao (1981), *Chi-square Tests for Homogeneity with Proportions Estimated from Survey Data*, New York: Academic Press.

Singer, E. (2006), "Nonresponse Bias in Household Surveys",*Public Opinion Quarterly*, 70(5), pp. 637–645.

Tourangeau R., G. G. Frederick, and M. P. Couper (2013), *The Science of Web Surveys*, Oxford: Oxford University Press.

Tourangeau R. (2017), "Paradoxes of Nonresponse", *Public Opinion Quarterly*, 81, pp. 803-814.

Williams D., and M. J. Brick (2018), "Trends in U.S. Face-to-Face Household Survey Nonresponse and Level of Effort", *Journal of Survey Statistics and Methodology*, 6, pp. 186-211.